INTRAURBAN VARIATION IN HOUSE PRICE APPRECIATION: A CASE STUDY, JACKSONVILLE, FLORIDA, 1980-1990

By

GREG T. SMERSH

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1995

This dissertation is dedicated to the memory of Martha E. Smersh (1925 -1987)

ACKNOWLEDGMENTS

A great deal of gratitude is due to all committee members for their generous help and support. More than anyone, Dr. Timothy Fik has offered countless hours of advice and direction and has given a substantial amount of inspiration to this work. Dr. David Ling has also provided much support and motivation. Additionally, other committee members, Dr. Edward Malecki, Dr. Peter Waylen, and Dr. John Dunkle, have offered helpful guidance and suggestions; their help and encouragement is most appreciated.

TABLE OF CONTENTS

APPENDICES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

INTRAURBAN VARIATION IN
HOUSE PRICE APPRECIATION:
A CASE STUDY,
JACKSONVILLE, FLORIDA, 1980-1990

By

Greg T. Smersh

August, 1995

Chairperson: Timothy J. Fik
Major Department: Geography

Classic land use and location theory suggest that residential property values in an urban area and the temporal changes in those values will vary spatially. However, there is a lack of research that defines how location affects land and housing prices and virtually no investigation of the spatial variation in house price appreciation.

Within an urban area, it is reasonable to assume that structural prices are spatially constant, at least from a cost perspective. Similarly, economic depreciation--the effect of age on a structure--can be assumed to be spatially constant. However, the assumption that intraurban land prices, or changes in those prices, are spatially constant may be considered naive.

This study investigates the Jacksonville, Florida, housing market from 1980 to 1990. Two hedonic housing price models are compared: a "naive" aspatial model and a

spatial model that includes location as an "interactive" component. The latter framework incorporates a theoretical accessibility function that denotes the unit price of land at different locations in space. As a general specification, a polynomial expansion of the function f(X,Y) over a Cartesian (X,Y) coordinate system is employed. According to this specification, prices of structural characteristics are considered spatially constant while the price of land is allowed to vary spatially. The naive and interactive models are estimated for both single and multiple time periods. With respect to coefficient estimation and spatially dependent error terms, the interactive, simultaneously estimated model is shown to be the superior specification.

Differences (percentage change) between price model equations for adjacent time periods are used to predict house price appreciation over space using a standard bundle of housing characteristics. Appreciation is then evaluated as a function of housing characteristics and location. While housing characteristics seem to have a negligible effect on appreciation, a definite spatial pattern emerges; this supports the notion that location plays an important role in house price appreciation. A repeat-sales methodology is employed to verify the existence of the implied positive and negative abnormal appreciation. In addition, the methodology estimates the geographic extent (radial distance from a maximum or minimum) of aberrant appreciation.

CHAPTER 1
INTRODUCTION

The importance of location as a determinant of property values has long been recognized in the theoretical literature on urban property valuation. Indeed, much theoretical and empirical work has been done studying the spatial variation of land and housing prices in the urban economics, geography, and real estate literatures. Similarly, the price appreciation (or returns) of real property with respect to time (but not space) has been examined in the economics, finance, and real estate literatures. However, there has been very little investigation of the spatial variation in house price appreciation.

House price appreciation is important to U.S. homeowners whose wealth is typically dominated by home equity. If appreciation varies spatially, as some theories suggest, then certainly such knowledge should be of interest to owners of both owner-occupied and renter-occupied (investment) housing. A more complete understanding of house price appreciation over space is also of importance to home mortgage investors and property accessors. Additionally, studies of the intraurban house price appreciation may be of interest to those who formulate housing policy, at least on a local (city or county) level.

The primary focus of this research is an investigation of alternative methodologies that could certainly be applied to any urban housing market, and more broadly, to any real estate (such as commercial) market. As an empirical case study however, this

research explores house price transactions in the geographical study site of Jacksonville, Florida during the time period of 1980 to 1990.

Jacksonville is located in northeast Florida on the Atlantic Ocean and encompasses all of Duval county, thus making city and county one and the same. Duval county and the urban area of Jacksonville are divided by the St. Johns River, the state's largest waterway. Jacksonville is one of the eastern seaboard's busiest deep-water ports, serving the U. S. Navy and merchant shipping. Jacksonville is also a major insurance and banking center. Two interstate highways--I-95 and I-10--run through downtown, linking Jacksonville to Los Angeles on the east-west axis, and to Maine and Miami on the north-south axis. Within an eight-hour drive on these highways live over 30 million people.

In the past 50 years, the state of Florida has seen tremendous growth, from a population of under 2 million in 1940 to nearly 13 million in 1990. During the 1980s, Florida saw a 30 percent increase in population, the second-highest increase of all states in the United States (Fik, Malecki, and Amey, 1993). Jacksonville (Duval county) has a 1990 population of approximately 672,900 people, up 18.5 percent from a 1980 population of 567,600 people.

During the 1980s, total employment in Florida rose over 40 percent, the highest percentage increase of all states in the United States. However, manufacturing employment in Florida increased only 16 percent during the same time period; in Duval county, manufacturing employment increased a mere 6.7 percent. However, per capita personal income during the 1980s rose 90 percent in the state of Florida and over 94 percent in Duval county.

Nationally, the median price of a single-family home rose from $62,200 to $95,500 in 1990, a 53.5 percent increase. In Duval county, house prices are much less than the national median but prices increased more during the 1980s. The median price of a single-family house in Duval county rose from $39,200 in 1980 to $62,700 in 1990, an increase of 60 percent. During this time period, interest rates on a 30-year fixed rate conventional mortgage fell rose from 13.77 percent in 1980 to a high of 16.63 percent in 1981 and then steadily fell to a low of 10.13 percent in 1990.

The national economy saw strong growth in the 1980s; gross domestic product (GDP) rose over 100 percent from 2.7 trillion in 1980 to 5.5 trillion in 1990. In response to this and the dramatic drop in both inflation and interest rates, the value of the U. S. securities market tripled between 1980 and 1990. It is important to point out, however, that these changes in the state and national economies would not be expected to have any influence on the intraurban variation of house prices in Jacksonville. Even local changes in population and employment, to the extent that they are spatially uniform, would not necessarily be expected to affect house prices.

Land use and location theory suggest that residential property values in an urban area will vary spatially and that intraurban variation of house price appreciation is also to be expected. Such theories are discussed in chapter 2, in a review of the literature. However, there is a lack of research that properly defines the extent to which location affects housing prices within the urban area and virtually no investigation of the spatial variation in house price appreciation. As the urban land and housing markets are so diverse and heterogeneous, the study of price variation over both time and space is a

difficult task. A review of various modeling methodologies and various specifications of location in house price models is also discussed in chapter 2.

Additional information on Jacksonville and the data employed in this research are discussed in chapter 3 while the methodologies employed in this investigation are discussed in chapter 4. This research first explores linear regression methods to measure the spatial variation of intraurban house prices. As casual observation indicates, the price of land varies spatially; the specification of land prices as spatially constant in a model may be considered, therefore, "naive." A naive model is compared to an alternative model which holds structural characteristics spatially constant but allows the price of land to vary spatially. Specifically, the alternative model incorporates a polynomial expansion of (X,Y) coordinates as a measure of the value of location (the unit value of land) at different locations in space and hence is presented as an "interactive" model. The naive and interactive models are estimated for single time periods and simultaneously over time utilizing single-equation and simultaneous-equation techniques. The objective of this investigation is to determine which model is best suited to separate the value of land from the value of the structure.

Results of the estimated house price models are then discussed in chapter 5. Here, the superior model specification is identified and used to predict land price surfaces. The price model uses a standard bundle of housing characteristics to predict prices for different time periods.

Chapter 6 discusses appreciation results. From the price equations, changes in the prices of land and structure are analyzed. Appreciation of structural characteristics is

compared to the appreciation of land alone and a composite price index of land and structure is compared to other price indices.

Areas of predicted "abnormal" appreciation are then identified. Abnormal appreciation is defined here as appreciation above or below two standard deviations from the mean rate of appreciation in the market. The existence of abnormal appreciation does not necessarily imply a spatial pattern; that consideration is next investigated by analyzing appreciation as a function of housing (structural) characteristics and location.

A repeat-sales technique is used to verify the existence of abnormal appreciation. Employing a spline regression procedure, the repeat-sales model is used to estimate the radial distances at which houses within exhibit the greatest difference in appreciation from the rest of the market. A summary and conclusion is presented in chapter 7.

CHAPTER 2
REVIEW OF THE LITERATURE

Review of Theory and Modeling

In this section, value theory and its unique application to land and housing is reviewed. Theoretical bases of the hedonic approach to modeling house values are also discussed. Finally, the theoretical aspects of house price appreciation, including the implications from value theory, are reviewed.

Land Value Theory

The value of a particular good, including land, is explained in various microeconomic value theories. Classical economists, such as Adam Smith (1776), viewed land value as a function of labor (as a factor of production) and recognized the income to land as a residual effect. As materials costs were fixed, labor was the integral component of production, and it was the cost of labor that determined the value of production. The labor cost premise was carried over to explain land income and value since land was considered a factor of production. However, the other factors of production, labor and capital, were mobile and could flow to locations that might provide greater returns. Therefore, labor had precedence over land for achieving a return, and land was considered to be price-determined by labor. The greater the marginal

productivity of a parcel of land, the greater the residual it provided to owners. This residual has been referred to as "surplus rent."

The German economist, Johann Heinrich von Thünen (1826), made a major contribution to land value theory by adding the element of location to marginal productivity. Von Thünen was concerned with the arrangement of different agricultural uses around a single market center. He theorized that the pattern of land use which developed was the result of different transportation costs (for each crop) and the intensity with which it was grown. He developed the concept of rent gradients for different agricultural land uses where rent is a function of the yield (or profitability) of a land parcel and, more importantly, the parcel's distance from the market.

Whereas classical theory places its emphasis on the cost of production (supply), marginal utility theory focuses on utility (demand). According to marginal utility theorists such as von Bohm-Bawerk (1888), the utility produced by the last unit of an economic good determines its value. Value is determined without consideration for costs of production; the short-run resolutions of marginal utility alone govern the theory. Alfred Marshall (1920) combined classical theory with marginal utility theory in his neoclassical market equilibrium theory, emphasizing that the interaction of both of these forces is important in the determination of value.

Monocentric City Models

Von Thünen's (1826) original concept of an agricultural monocentric model was generalized and applied to housing many years later by Alonso (1964). Models that

assume a monocentric city represent a unique branch of microeconomic theory; these models expand consumer behavior theory to incorporate the consumption of land and locational preference. The spatial factor complicates neoclassical economic theory because households must locate in only one location and no two households can occupy the same location. To simplify this problem, monocentric models assume that all employment is centrally located, that locational choice depends only on commuting costs and land consumption, and that housing capital is infinitely divisible and mobile.

Alonso (1964) assumed production and consumption decisions determined land consumption by households. In his model, the direct household preference for land determines residential density. Muth (1968), and later Mills, (1972) expanded the monocentric model to incorporate housing. In the Muth-Mills model, consumer utility depends on the consumption of other goods and an aggregate commodity, "housing." In the Muth-Mills approach, residential density is determined by the production function for housing. The major predictions of the monocentric models are that residential densities decline (at a decreasing rate) with distance from the central business district (CBD) and that house prices also decline with distance at a decreasing rate.

The basic assumptions of monocentric city models are unrealistic. In particular, housing capital is lumpy in size and nontransportable and locational decision making is not typically based on a trade-off between land consumption and commuting costs. Furthermore, while the form of many urban areas has tended towards a pattern of central employment in the past, the general pattern of urban employment in the contemporary city is much more dispersed. Few metropolitan areas have a single dominant node such as

the CBD, service employment is widely dispersed and there has been a decentralization of office and industrial establishments as well.

Despite questionable assumptions, many insights into urban housing markets have been derived from the works of Alonso, Muth, and Mills. The most intriguing observation is that housing and accessibility are jointly purchased. As Muth (1968) notes,

> until quite recently, most writings on urban residential land and housing markets tended to neglect accessibility. They emphasized instead the dynamic effects of a city's past development upon current conditions, and the preferences of different households for housing in different locations. (pg. 300)

The classical literature suggests that increases in the centrality (accessibility) of a parcel of land will generally lead to an increase in value. In other words, accessibility advantages due to location are capitalized in the price of housing.

Hedonic House Price Models

New modeling approaches were developed in the 1960s as a method to better understand the relationship between housing market prices and the components of "housing services" imbedded within them; these became known as "hedonic" models. A simplification of the heterogeneous aspects of urban housing stock was first accomplished by couching the demand for housing in terms of these housing services or "bundles" of housing attributes to estimate implicit characteristic prices. In this perspective, housing value is viewed as a bundle of (utility producing) services offered by a combination of structural and locational characteristics, the component prices of which are never directly observed in property transactions.

The interest in applying these methods to housing markets evolved from Lancaster's (1966) consumer theory of differentiated products; this theory proposed that all households have demands for underlying characteristics (inherent in all traded commodities) and that households combine these characteristics to produce "satisfactions." Focusing on the use of multi-variate models, hedonic studies aimed at uncovering consumer preferences for (structural) housing characteristics.

The hedonic (or preference) approach was also applied to estimating the effect of location and the impact of accessibility (primarily to employment centers), environmental amenities, and externalities. Such models employed distance gradients (such as miles from the CBD or an externality), gravity model expressions of accessibility, or dummy variables (for location in specific areas). The advantage of the hedonic approach is that it allows for the estimation of coefficients for each characteristic holding the effect of all others constant. Detailed discussions of the mechanics of hedonic price models are offered by Rosen (1974) and Little (1976).

Theory suggests that the value of land is a phenomenon that results from the forces of supply and demand. In turn, supply and demand are the market effects of the relative scarcity and utility associated with urban land. Transaction prices reflect supply and demand conditions and the outcome of a market-clearing process by which households of various incomes arrange themselves by geographic location and type of housing stock. Thus, the estimation of implicit prices represents not demand but rather an estimate of the (upper) bid-rent function of different buyers for particular housing components and the (lower) offer function of different sellers.

In an early regression model, Brigham (1965) sought to ascertain determinants of residential land value. This study utilized data on land value gradients (measured in price per square foot) along three vectors which extended from the city (CBD) center of Los Angeles to Los Angeles County boundaries. Brigham suggested that land value was a function of a site's accessibility, amenity level, topography, and certain historical factors that affect its utilization. As Brigham observes,

> urban land has a value over and above its value in rural uses because it affords relatively easy access to various necessary or desirable activities. If transportation were instantaneous and costless, then the urban population could spread out over all usable and all land prices would be reduced to their approximate value in the best alternative use. (pg. 326)

Brigham created an accessibility potential variable that measured the accessibility potential of each site to multiple workplaces; other variables included distance to the CBD, an amenity variable (average neighborhood house price), and a topography dummy variable. Regression equations were fitted to spatial moving averages of the value per square foot for single family properties on each vector. The data were smoothed in this manner to remove as much spurious variation as possible and to allow the investigation of general, not local, variations in land values. This empirical investigation provided strong statistical evidence to support the concept of property values as a function of structural and neighborhood characteristics and accessibility to employment.

Other researchers have measured accessibility in similar ways. An investigation of land values in Topeka, Kansas by Knos (1968) compared linear and nonlinear gradients (of distance to the CBD) to a generalized (population potential) accessibility index, also derived from a gravity model. Alone, the index was only marginally significant; however,

its combination with distance gradients provided a highly significant model. With regard to the importance of (CBD) workplace accessibility, the empirical evidence is somewhat mixed. In Kain and Quigley's (1970) study of the St. Louis housing market, the inclusion of a distance variable (in miles from the CBD) was found to be statistically insignificant. However, the works of Brigham (1965) and Knos (1968) suggest that such a finding may simply be the result of model misspecification.

Other early developments in hedonic price models tended to view property price solely as an additive property of hedonic characteristics (Berry, 1976, and Linneman, 1980, and 1981). Berry (1976) and Berry and Bednarz (1977), investigated price differences in ethnically distinct housing markets in Chicago. These analyses sought to study market segmentation based on race and income; specifically, they concluded that single-family housing prices in Black and Hispanic neighborhoods were significantly less than in White neighborhoods.

Henderson (1977) suggests that the external benefits or costs of particular land uses or urban activities will be capitalized into property values. With regard to such "externalities" or spillover effects, analyses of hedonic prices have provided direct evidence of residential blight (Kain and Quigley, 1970); air pollution (Anderson and Croker, 1971; Harrison and Rubinfeld, 1978); closeness to appealing amenities (Weicher and Zerbst, 1973); neighborhood characteristics (Berry and Bednarz, 1977); proximity to non-residential land use (Li and Brown, 1980); nearness to a potentially dangerous land use (Balkin and McDonald, 1981); environmental amenities (Gillard, 1981) and proximity to waste disposal (Thayer et al., 1992), among others.

While the estimated coefficients of such measures in hedonic regression are usually significant, Ball (1973) notes that the independent effect of distance (or generalized accessibility) is often rather small. In part, this may simply reflect the negative covariance between accessibility and housing vintage--the tendency in most U.S. cities is for older, more obsolete housing units to be located closer to traditional central employment centers.

Hedonic regression models have often included neighborhood externalities with dummy variables (for location in specific zones) or distance gradients. However, such models have generally underspecified the locational characteristics of housing as they have not included the influences of all urban nodes (employment centers, schools, shopping centers, etc.), axes (highways and major arterials), and externalities (parks, landfills, airports, etc.).

House Price Appreciation

Several theories suggest that the temporal change or appreciation in house prices (ex-ante) will vary within an urban area. First, Muth (1975) has demonstrated that rising real income and population have caused net (implicit) rental income (and therefore housing prices) to increase faster at the city fringe than at the city center. Second, various studies of house prices and rental income indicate that housing depreciates at a decreasing rate with the age of the unit. All else the same, this should produce varying rates of appreciation among submarkets as the vintage of the housing stock is not uniform across the metropolitan area (Archer, Gatzlaff, and Ling, 1995).

Further, deLeeuw and Struyk (1975) demonstrate a "filtering model" which indicates that larger houses will experience more rapid price appreciation. The demand for housing has been shown to be income elastic and therefore, rising real income in an urban area tends to generate an increased demand for larger houses and a corresponding decrease in demand for smaller, less functional houses. Housing unit size clearly exhibits spatial variation although such variation is more likely to be scattered and have less of a spatial pattern than housing age or lot size.

Finally, the theoretical models of land price may indicate foundations for theories of land price appreciation. The theoretical and empirical literatures suggest that increases in accessibility will lead to increases in property value. If accessibility is interpreted in a general connotation of the word (accessibility to work, accessibility to shopping, accessibility to crime, accessibility to appealing amenities, etc.), then, in aggregate, accessibility defines the location of a specific site. Any changes in accessibility benefits (or dis-benefits) may be due literally to increased access (a new road) or simply an increase/decrease in an activity (shopping, crime). Thus, theory may imply that such changes in accessibility advantages over time will be reflected in changes in price.

In addition to the ex-ante effects of perimeter location, house age, and house size, ex-post appreciation may also be affected by unanticipated changes in the value of housing's physical or locational characteristics. For example, localized storm damage may result in significant price changes; such unanticipated exogenous "shocks" may increase or decrease prices dramatically, especially over short time periods. The effects of other events such as the construction of a new highway or shopping mall may be

significant over longer time periods; such events may be seen as significantly changing accessibility benefits in an urban area.

The limited empirical evidence available does suggest that house price appreciation is affected by location within the urban area. Using hedonic techniques and five metropolitan areas, Rachlis and Yezer (1985) find that the rate of change in house prices is statistically related to location characteristics of housing. As their measures of location, Rachlis and Yezer (1985) used distance to the CBD, distance to a high income neighborhood, and distance to a minority neighborhood. Keil and Carson (1990) find a statistically significant difference in appreciation between incorporated and non-incorporated locations within a metropolitan area.

Defining neighborhoods by zip codes, Case and Shiller (1994) find that property values in Boston and Los Angeles appreciate at similar rates when the metropolitan area as a whole is performing well. However, they find substantially more dispersion in appreciation when the metropolitan area is experiencing price declines.

Using a repeat-sales methodology in a cross sectional study of Miami census tract groups over a 22-year time period, Archer, Gatzlaff, and Ling (1995) seek to determine if there is significant locational variation in house price appreciation and find that over half of the 79 tract groups show statistically significant abnormal (annual) appreciation. The repeat-sales methodology uses only houses which have sold twice during a specific time period to generate an overall price index or sets of indices for different areas. Their procedure generates a pair of indices that compare each tract group to the combination of all other tract groups; the process is repeated for all 79 tract groups. Abnormal

appreciation is defined (for a census tract group) as a rate of appreciation that is significantly different from the rest of the market. However, tract group location explains only 12 percent of the variation in appreciation that is unexplained by market-wide price movements. Abnormal appreciation here appears to be dominated by influences at the individual house level or perhaps an alternative (i.e., smaller) geographic level.

Theoretical Summary

The contributions of economic theory to the perception of value conclude that value is a market concept. Marshallian theory--the neoclassical approach--integrates all other relevant theories into the supply-demand model. Supply and demand are the market effects of the relative scarcity and utility associated with a particular good.

Land as an economic good complicates neoclassical economic theory because households must locate in only one location and no two households can occupy the same location. Theoretical models, such as the monocentric city models of Alonso, Muth, and Mills must therefore impose strict assumptions to simplify the situation. While these assumptions are quite unrealistic, many insights about the interrelationships of urban housing markets have come from the observation that housing and location are a composite or "bundled" good.

Hedonic models seek to uncover consumer preference (or utility) for different components of housing which are never directly observed in actual property transactions. Hedonic models can be used to differentiate various housing characteristics, including location. A review of alternative methodologies is provided in the following section.

Review of Alternative Methodologies

The previous section discusses hedonic models as a methodology for separating the value of various housing characteristics. Typically, spatial effects are derived from hedonic models at a given point in time, while temporal effects are estimated using either hedonic or repeat-sales methods that may include various measures of location. This section first discusses the use of rudimentary hedonic equations for price indices and the derivation of the repeat-sales technique. The remainder of the chapter discusses methods for incorporating various measures of location in more complex hedonic equations.

Price Equations and Indices

The basic hedonic house price model regresses transaction price on structural characteristics (such as square footage and age), land characteristics (such as lot size), and locational (or neighborhood) characteristics. This approach can be used to generate a temporal price index in several ways. Alternatively, house price indices can be generated using data on only those houses which sold twice--the repeat-sales technique.

The advantage of the repeat-sales technique is that it avoids the temporal variation in characteristic prices manifest in hedonic estimation; significant variation in these prices may bias index results. This technique is derived as the difference between two hedonic equations for different time periods; constant quality (no change in housing attributes over time) is assumed and so hedonic variables drop out of the estimating equation, leaving only time as an explanatory variable.

Hedonic Price Index Estimation

Generating the hedonic index requires a sample of house sales from multiple time periods. Transaction prices are regressed on structural and locational characteristics. Once the hedonic equation has been estimated, it can be used to produce a price index. There are two major models: "strictly cross-sectional" and "explicit time-variable."

In the strictly cross-sectional model of house prices, the implicit characteristic prices are estimated in a separate hedonic regression for each time period, thereby allowing the implicit characteristic prices to vary over time. A model of the following type is common (e.g., Berry, 1976):

$$P_i = \beta_0 + \Sigma_j^k \beta_j X_{ji} + \varepsilon_i \qquad (1)$$

where P_i is the transaction price of property i, i = 1 to n observations, and β_j denotes a vector of coefficients, j = 1 to k, on the structural and locational attributes, X_{ji}, which could include square footage, age, lot size, and various neighborhood characteristics. The coefficient β_0 is an intercept term and ε_i is a random, normal, independent error term.

Price indices are then predicted for each period by applying the estimated implicit prices to a standardized bundle of housing attributes. This model is often used in a single time period when measuring spatial effects. With time held constant, location in space can be measured in a more distinct manner; the simultaneous estimation of price over both time and space is more complicated.

The explicit time-variable approach includes time as an independent dichotomous variable; the following is a popular functional form (e.g., Clapp and Giaccotto, 1991):

$$\ln P_{it} = \Sigma_j^k \beta_j \ln X_{jit} + \Sigma_t^T c_t D_{it} + \varepsilon_{it} \tag{2}$$

where "ln" denotes natural logarithm, P_{it} is the transaction price of property i at time t, t = 1 to T time periods, and β_j indicates a vector of coefficients on the structural and locational attributes. Here, c_t denotes a vector of time coefficients on D_{it}, time dummies with values of 1 if the house sold in period t and 0 otherwise. From this equation, the anti-logarithm (e^x) of the coefficient c_t, scaled by 100, then becomes a (cumulative) price appreciation index. This model is discussed by Clapp and Giaccotto (1991) and Gatzlaff and Ling (1994). Potential problems associated with the hedonic technique, including model misspecification, sample selectivity and the choice of functional form, as discussed by Palmquist (1980) and Halvorsen and Pollakowski (1981). These problems can be partially overcome by employing the repeat-sales technique.

The Repeat-Sales Technique

The repeat-sales technique allows for the estimation of intertemporal market price indices for "quality-adjusted" or standardized properties. The origins of this technique can be traced back to the work of Bailey, Muth, and Norse (1963) and are discussed by Hall (1971), Palmquist (1980), Case (1986), and Gatzlaff and Ling (1994). This technique is a modification of the explicit time variable approach that uses a chain of overlapping time periods to predict cumulative appreciation rates for specific time periods.

More precisely, the repeat-sales model is the difference between the log of a "second" sale model and the log of a "first" sale model. From equation (2) then

$$\ln P_{it} - \ln P_{i\tau} = \left(\Sigma_j^k \beta_j \ln X_{jit} + \Sigma_t^T c_t D_{it}\right) - \left(\Sigma_j^k \beta_j \ln X_{ji\tau} + \Sigma_\tau^T c_\tau D_{i\tau}\right) + \varepsilon_{it} \tag{3}$$

where P_{it} and $P_{i\tau}$ are the prices of repeat-sales transactions, with the initial sale at time τ and the second sale at time t for t = 1 to T time periods. If housing quality is constant, (the implicit assumption in the repeat-sales technique) then structural and locational variables cancel out and the difference between the two prices is solely a function of the intervening time period. Under this condition, equation (3) reduces to

$$\ln P_{it} - \ln P_{i\tau} = \Sigma_t^T c_t D_{it} - \Sigma_\tau^T c_\tau D_{i\tau} + \varepsilon_{it} \tag{4}$$

To execute this procedure, the dependent variable is the log of the price ratio generated from a property having sold twice. The log of the price ratio is then regressed on a set of dummy variables, one for each period in the study. The repeat-sales estimating equation is

$$\ln (P_{it} / P_{i\tau}) = \Sigma_t^T c_t D_{it} + \varepsilon_{it} \tag{5}$$

where $P_{it} / P_{i\tau}$ is the ratio of sales price for property i in time periods τ and t; D_{it} is a dummy variable which equals -1 at the time of initial sale, +1 at the time of second sale,

and 0 otherwise; and c_t is the logarithm of the cumulative price index in period t. To clarify, $c_t = \ln(1 + A_t)$, where A_t is the cumulative appreciation rate for year t.

The repeat-sales model avoids many of the problems associated with hedonic models, but is subject to several criticisms. Case and Shiller (1987) and Haurin and Hendershott (1991) note that the sample may not be representative of the housing stock, upgrading of the property may be ignored, and attribute prices may change over time.

Given the somewhat restrictive functional form of the basic repeat-sales model, any measures of location must properly be included as interactive (as opposed to additive) terms; the inclusion of one location (dummy) variable will double the number of estimated coefficients. With respect to the generation of temporal indices, Gatzlaff and Ling (1994) find that the "strictly cross-sectional" models and "explicit time-variable" hedonic models with limited variables of square footage, age, and lot size produce indices similar to those estimated with repeat-sales.

Multinodal Models

To be relevant today, the monocentric model must be extended to represent the modern urban setting and recent research has sought to incorporate additional measures of location within a multinodal context. This results in multiple price gradients and may undermine the significance of the CBD as a single influence. Heikkila et al. (1989) present a model of residential land values which explicitly incorporates distance from multiple employment centers. The conclusion of this study of housing in Los Angeles County is that the CBD price gradient becomes statistically insignificant once distances

to employment centers are included. This finding contradicts one of the principle features of the monocentric model.

Point pattern analyses by Green (1980) and Getis (1983) influenced the hedonic price models of Waddell, Berry, and Hoch (1993), which explicitly incorporate distance from multiple market (or employment) centers. Their investigation of the Dallas housing market examined the implicit price of relative location over discrete measures of distance (rather than continuous gradients) in a multi-nodal area.

Waddell, Berry, and Hoch included both temporal and spatial effects but did not allow measurement of an interactive effect; the model form extends from equation (2):

$$\ln P_{it} = \Sigma_t^T c_t D_{it} + \Sigma_j^k \beta_j X_{jit} + \Sigma_m^k \lambda_m D_{mit} + \varepsilon_{it} \tag{6}$$

where P_{it} is the transaction price of property i at time t; c_t denotes a vector of time coefficients of D_{it}, time dummies with values of 1 if the house sold in period t and 0 otherwise. Here, β_j denotes a vector of coefficients on the structural and locational attributes, X_{jit}, such as age of construction, wall type, log of living area, and percent of land in census tract for various land uses. As measures of relative location, λ_m denotes a vector of coefficients of D_{mit}, dummy variables based on distance intervals of less than one mile, one to two miles, two to five miles, and five to ten miles from major urban nodes.

Although equation (6) allows for the creation of a house price index that includes spatial effects, it specifies space as discrete rather than continuous and assumes that time

and space have additive effects on property price because both space and time are represented with dummy variables. The model is easy to interpret; however it does not consider any interactive effects of space and time. Therefore, the model does not properly measure price appreciation over space.

Heikkila et al. (1989) and Waddell, Berry, and Hoch (1993a, 1993b) incorporate such explanatory variables as accessibility to suburban employment centers, expressways, and other nodes and axes of influence. Waddell, Berry, and Hoch (1993a) find that,

> the emergence of new nodes of regional significance has created house price gradients that far overshadow any residual gradient with respect to the CBD. Moreover, the raw price gradients surrounding these new nodes are almost completely explained by structural and neighborhood variables, indicating the degree to which the physical stock and the form of neighborhood externalities have been reshaped in response to these emergent spheres of influence. In older established areas of the city it has been much more difficult to adjust the housing stock, and both depreciation and negative externalities far outweigh residual price-distance gradients. (pg. 15)

Although empirical evidence has supported the theory that the land value gradient declines with increasing distance from central points within an urban area, it is the work of Johnson and Ragas (1987) that examines the spatial influence of externalities within the CBD. They contend that it is centrality (accessibility in general) and multiple externalities that influence land values. Johnson and Ragas (1987) develop a model for undeveloped urban land and explore various model specifications and functional forms using data from New Orleans. From equation (1), but including time and distance variables, Johnson and Ragas estimate

$$P_{it} = \Sigma_t^T c_t D_{it} + \Sigma_j^k \beta_j X_{jit} + \Sigma_m^k \lambda_m R_{mit} + \varepsilon_{it} \tag{7}$$

where P_{it} is the transaction price per square foot of property i at time t, i = 1 to n, and t = 1 to T; and c_t denotes a vector of time coefficients of D_{it}, time dummies with values of 1 if the house sold in period t and 0 otherwise. Here, β_j denotes a vector of coefficients on the spatial and aspatial plot-specific characteristics. As measures of relative location, λ_m denotes a vector of coefficients of R_{mit}, distances from positive and negative externalities.

An expanded model considers the interactive effect between X_{jit} (zoning) and R_{mit} (distance). Alternative functional forms in addition to the linear model were also estimated including a log-linear transformation (of P_{it}), and a Box-Cox transformation.

Trend Surface Analysis

Johnson and Ragas (1987) then compare their (price gradient) models to trend surface analysis (TSA) models. They find that the TSA models better predicts land prices (based on values of R^2). TSA offers a way to measure price variations in a purely spatial context.

TSA is a technique of fitting (absolute) spatial data by regressing the variable in question (such as land value) on a p^{th} order polynomial expansion of the Cartesian coordinates for each data value (Hembd and Infanger, 1981; Parker, 1981). The general form of the absolute location or trend surface (TSA) model used by Johnson and Ragas (1987) is

$$P_i = \Sigma_j^p \Sigma_k^p \beta_{jk} [X_i^j Y_i^k] + \varepsilon_i \qquad (8)$$

where P_i is the price per square foot of property i; β_{jk} denotes a vector of coefficients of X_{ij} and Y_{ik}, Cartesian coordinates of the properties in the sample and $j + k \leq p$, where the model is a p^{th} order polynomial.

Trend surface mapping has traditionally been used in engineering and the geological sciences (Krumbein and Graybill, 1965). TSA applications to geographical research are presented by Chorley and Haggett (1965). Although the TSA price equation lacks any explanatory meaning and the only way to demonstrate model results is visually--the comparison of this pure spatial model to graphic displays of other (behavioral) hedonic models can provide valuable insight.

A trend surface analysis of property values throughout an urban area demonstrates how urban spatial structure affects (localized) price gradients. TSA not only identifies prominent nodes on the landscape, it also shows the value at those nodes, the slope of the price (value) gradient, and thus the effect of proximity to a node. However, the trend surfaces would be expected to vary tremendously for different land uses. For example, the demand for accessibility to retail sites is much better defined than the demand for accessibility to (from) residential sites. The spatial variation in house prices is often so great that the observation of spatial patterns in individual prices is difficult and areal aggregation may become necessary.

Accessibility Indices

The models of Alonso (1964), Muth (1969), and Mills (1972) suggest that increases in the accessibility of a parcel of land in an urban area will generally lead to

an increase in the value of that parcel. The hypothesis that accessibility plays a prominent role in the determination of house price and house price appreciation suggests that researchers would be intent on determining if spatial variations are observable. However, there is a lack of research that properly defines relative location (or a general accessibility index) in such a manner as to capture all of the multinodal features of the urban landscape.

Despite the importance of location, few hedonic price equations have been constructed to include more sophisticated measures of accessibility; a notable exception was the contribution of Jackson (1979). In a study of the Milwaukee housing market, Jackson uses house rents from the U. S. Census Bureau for one time period at the (census tract) geographic level to derive a continuous measure of house price (rents) over space. What is most significant about this model is its capacity to isolate the influence of location or accessibility in general on the price of housing in the following form, extending from equation (1):

$$P_i = \beta_0 + \Sigma_j^k \beta_j X_{ji} + \Phi (A_i)L_i + \varepsilon_i \tag{9}$$

where price (in this case, census tract rent), P_i, is represented as a linear function of a constant β_0, a vector of variables which define structural and neighborhood characteristics (X_{ji}), and the quantity of land (L), as measured by lot size. The coefficients β_j represent a vector of structural and neighborhood characteristics, and Φ(A) is the price of land as a proxy for accessibility.

The theoretical accessibility function $A = f(X_j, Y_k)$ denotes the level of accessibility at location (X_j, Y_k) using Cartesian coordinates X_j and Y_k. If the function f were known, the level of accessibility at a given location could be evaluated with respect to the spatial distribution of all prominent nodes (employment centers, retail shopping outlets, schools, etc.) As a general specification, Jackson (1979) employed a Taylor series expansion of the function f(X,Y) about the midpoint of a Cartesian coordinate system, yielding:

$$A_i = \Sigma_j^p \Sigma_k^p a_{jk} [X_i^j Y_i^k] + r_i \qquad (10)$$

where r_i is a remainder and $j + k \leq p$. Although equation (9) is written in p^{th} order polynomial form, a remainder exists to account for the inexactness of the transformation at order p. Equation (9) is a representation of a double power series formula, equation (8), that is widely used in trend surface analysis. Substituting equation (10) into equation (9), the underlying dependence of land value on accessibility produces "a double power series representation of land price." According to this model specification, hedonic prices of structural and neighborhood characteristics are considered spatially constant while the price of land varies spatially as a result of demand for more accessible sites. This model formulation "is consistent with theories of urban land value which hold that accessibility advantages are capitalized in the land price." (Jackson, 1979, pg. 467)

An OLS assumption that has a high potential to be violated and yet often goes unchecked in hedonic price equations which incorporate measures of location is that error terms are not spatially correlated; this problem is discussed by Cliff and Ord (1973).

However, Jackson's methodology for incorporating accessibility was also shown as a way in which to reduce the likelihood of encountering estimation problems caused by spatially dependent (autocorrelated) error terms.

Price Model Summary

From the alternative model specifications reviewed here, it is Jackson's (1979) polynomial expression of land prices that seems the most promising for the examination of spatial variation in house price appreciation. Jackson's work is the foundation for this research; however, in this research there are significant differences. Jackson used census tract rents for one time period while this investigation uses actual house sales (aggregated at a much smaller geographic level) to estimate house price equations for multiple time periods. These data are discussed in more detail in the following chapter while specific methodologies are discussed in chapter 4.

CHAPTER 3
DATA

Source and Scope of the Data

This research analyzes the Jacksonville, Florida housing market. With regard to boundaries, the city of Jacksonville is synonymous with Duval County. As this study is concerned with urban housing, a 154-square-mile study area (see Figure 3-1) is defined.

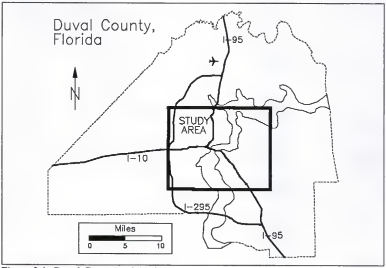

Figure 3-1 Duval County and Study Area

This (11 mile by 14 mile) study area contains what could be characterized as an urban density of housing. It is physically bounded to the west by Interstate 295 and to the north by the St. Johns River and is logically bounded in all directions by a paucity of housing. The study area and major urban nodes and axes are shown in Figure 3-2.

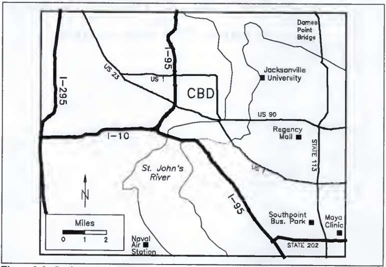

Figure 3-2 Study Area and Major Nodes and Axes

The data come from the Florida Department of Revenue's (DOR) property tax records. These data are compiled each year by the DOR and maintained as a multi-tape database which includes information on every parcel in the state of Florida. These data include square footage, age, lot size, last sales price and date, and previous sales price and date. To adjust for any mispricing (due to improvements, family sales, etc.) the data have

been carefully cleaned; detailed procurement, cleaning, and manipulation procedures are presented in Appendix A.

The data span the years 1979 to 1990 and are aggregated temporally into biannual time periods. Within the 154-square-mile study area, there are an average of 1,928 sales per biannual time period or a total of 11,570 sales over the entire 12 year period. Of these, there are 3,998 houses which sold twice and are used in the repeat-sales analysis.

GIS Procedures

Using a geographic information system (GIS) address matching procedure, all properties are geo-coded. This process searches a street database and interpolates a (latitude / longitude) point based on the house number contained within the range for its block. Latitude / longitude coordinates are then converted into Cartesian coordinates with an origin at the southwest corner of the county.

GIS is also used to determine optimal areal units (described below) and, using a point in polygon procedure, aggregate individual property characteristics into the specified areal units. Finally, once points of maximum (minimum) appreciation are identified, GIS is used to calculate distances from every house to those points; this is for use in the repeat-sales spline regression.

Aggregation Techniques

The data are aggregated both geographically and temporally. The rational for geographic aggregation is that too much "noise" exists at the individual house level; that

is, there is excess variation in house price beyond that which can be explained by square footage, age, and lot size. The rational for temporal aggregation is, first, that there are only minor price changes over space on an annual basis and, second, that temporal aggregation allows geographic aggregation at a smaller geographic level.

Geographic Aggregation

A number of preliminary tests using third and fourth order expansions of Jackson's (equation 9) model are performed to determine an optimal aggregation technique. Using individual sales, about half of the interactive terms are significant but the overall explanatory power of the model is lower (R^2 statistics of approximately 0.80) than expected. This is likely due to unobservable differences such as maintenance, overall quality, and amenities in individual houses.

Aggregation at the census tract level is too broad; the explanatory power of the model is improved (R^2 statistics of around 0.85) but few interactive terms are statistically significant. Aggregation at the census block group level produces better results (R^2 statistics nearing 0.90 with over half of the interactive terms significant) but the number of house sales vary tremendously between block groups.

A spatial moving average (using 1 mile radial areas at 1 mile increments) is also created; this produces superior results (R^2 statistics over 0.90 with most of the interactive terms highly significant). However, this method is rejected because of the double counting of house sales. Finally, a 140 cell grid (see Figure 3-3) system that seeks to minimize the variation in number of sales between geographic units is partitioned.

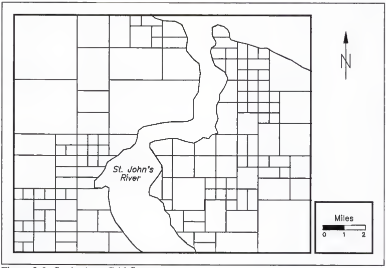

Figure 3-3 Study Area Grid System

This system is based on a grid of quarter sections where a section equals one square mile. Quarter sections are joined in such a manner that each unit contains as nearly an equal number of sales for each time period; there are an average of 14 sales per areal unit and a minimum of 4 sales per areal unit for each time period. This model has good explanatory power (R^2 statistics of 0.94 to 0.96) and many of the interactive terms are highly significant. It is the preferable method because it allows the greatest number of geographic units with the most evenly aggregated number of house sales. This grid system reflects a more even spatial distribution of house sales.

In aggregating the data, the mean is taken for the variables of price, square footage, age, and lot size for each grid cell for each time period. The (X,Y) coordinates

for each cell are not the cell centroid but, rather, the mean (X,Y) coordinate for all houses (regardless of year of sale) in that grid cell. These coordinates are then used in the interactive terms for all time periods.

Temporal Aggregation

The data are aggregated temporally into biannual time periods to allow geographic aggregation within a greater number of (smaller) grid cells. Additionally, it is observed that there is a relatively minor change (about 5 percent) in prices on an annual basis. However, strong motivation exists for the adjustment (compounding forward or discounting back) of house prices. Because models seek to estimate price changes over both time and space, a greater possibility exists for bias (due to time of sale) between geographic units.

A price index is created for the entire study area using average house sales; this index is nearly identical to indices created with hedonic regression and repeat-sales. These annual urban appreciation rates are used to adjust actual sales prices on a monthly basis. For example, time period "1980" contains 1979 sales which are compounded forward to the midpoint of the 24-month period and 1980 sales which are discounted back to the midpoint. That is, for time period 1980, individual sale prices are compounded / discounted to January 1, 1980 using the price index. In this fashion, data sets are created for the 6 biannual time periods, 1980 through 1990. These data sets will be used to estimate six strictly cross-sectional hedonic models; the equations are then used to predict appreciation rates over space.

Repeat-Sales Data

Data used in the repeat-sales technique are individual houses which sold twice; these data thus preserve information that is lost in aggregation. The spatial distribution of these (3998) data points is shown in Figure 3-4.

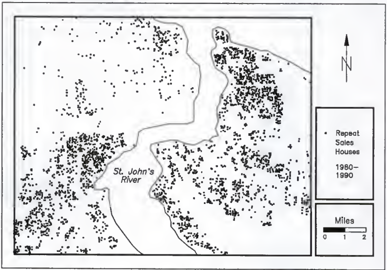

Figure 3-4 Spatial Distribution of Repeat Sales Observations

This data set is used to test the existence of any predicted abnormal appreciation. Additionally, it is used to estimate the radial extent of any abnormal appreciation; the methodology is discussed in the following chapter.

CHAPTER 4
METHODOLOGY

Hedonic Price Equations

A hedonic regression model that allows land prices to vary spatially is not only intuitively appealing, but may provide more accurate structural and locational coefficient estimates. Additionally, simultaneous-equation estimation may be preferable to single-equation estimation as contemporaneous correlation may be present in the error structure of the models. Regression assumptions are discussed in Appendix B.

Using the data discussed in the prior chapter, four model specifications are tested. These are the naive model with single-equation estimation; the naive model with simultaneous-equations estimation; the interactive model with single-equation estimation; and the interactive model with simultaneous-equations estimation.

Naive versus Interactive Model

Models that incorporate the price of land as either spatially variant or aspatial are compared here; the aspatial or "naive" model is the standard hedonic price equation defined in equation (1):

$$P_i = \beta_1 \text{ SQFT}_i + \beta_2 \text{ AGE}_i + \beta_3 \text{ LOT}_i + \varepsilon_i \qquad (11)$$

where P_i is the mean transaction price of all houses in grid cell i, i = 1 to 140, and estimated as a linear function of $SQFT_i$, the mean structural square footage, AGE_i, the mean building age, and LOT_i, the mean lot size. Following from Jackson's (1979) model, an interactive model that interacts lot size with a polynomial land price surface is derived:

$$P_i = \beta_1\ SQFT_i + \beta_2\ AGE_i + \Sigma_j^k\ \beta_j \left[LOT_i * f_{ij}(X,Y) \right] + \varepsilon_i \tag{12}$$

where $f_{ij}(X,Y)$, j = 1 to k, is a polynomial expansion of (X, Y) coordinates. According to this model specification, the structural characteristics of square footage and age are considered spatially constant while lot size interacts with the polynomial terms, allowing the price of land to vary spatially.

The primary advantages of this specification over Jackson's (1979) model are: the use of actual sales data; multiple time periods; aggregation at much smaller geographic units; and the origin of the Cartesian coordinate system. Although Jackson's use of an origin at the data (X,Y) median allows interactive coefficients to be interpreted as partial derivatives, there is no theoretical justification for his "double power series" representation of price that this method manifests. Based upon preliminary tests of both methodologies, an origin outside of the data set seems to offer more reasonable results and is used here.

Single versus Simultaneous Estimation

Separate estimation of the interactive model for each of the six time periods produces many coefficient estimates on square footage, age, and the interactive terms that

exhibit relatively strong temporal patterns. These temporal patterns suggest a spatial pattern of house price appreciation. If this is so, the data should be regarded not only as cross-sectional but as time-series as well. This implies that error terms in equations for different time periods may be autocorrelated at a given point in time but not necessarily correlated over time. This is known as contemporaneous correlation and is discussed by Judge (1985).

One method of combining cross-sectional and time-series data effectively "stacks" the regression equations and estimates model coefficients (for all time periods) simultaneously via a generalized least squares (GLS) technique. The possible gain in the efficiency of the model obtained by simultaneously estimating price equations for all time periods led Zellner (1968) to assign the title "a set of seemingly unrelated regression equations." Seemingly unrelated regression (SUR) estimation is employed for both the naive and interactive price models. A more technical description of the seemingly unrelated regression procedure is presented in Appendix C.

Spatial Autoregressive Variable

The interactive model will estimate an overall assessment of the intraurban variation in land price. To examine more "localized" effects, a spatial autoregressive price variable is created. This is an average of price in all contiguous grid cells; the variable is defined as follows:

$$SAP_i = \left[\Sigma_j^n c_{ij} P_j \right] / \Sigma_j^n c_{ij} \qquad (13)$$

where SAP_t is the spatial autoregressive price variable and c_{ij} is a binary connectivity matrix that denotes the connectivity of each cell with all other cells. The matrix is based on what Cliff and Ord (1973) refer to as a "Queen's case" (edge-to-edge and vertex-to-vertex) set of joins. This variable is created for each time period and tested as an additional variable in the interactive models.

Model Estimation

The results of the four model specifications are compared with respect to coefficient estimation and spatially dependent error terms. Based on these criteria, the "superior" model specification is selected.

Component Prices of Structure and Land

The estimated coefficient price for square footage represents the unit price of structural components; over time, this coefficient should more or less emulate a general construction price index. Because the cost of materials to construct a house will generally rise over time in an approximation of such an index. Estimated coefficient prices on square footage are compared to the Producer Price Index (PPI) for construction materials, a national index. These coefficients would be expected to be positive and can directly be interpreted as the price per square foot to construct a new house in the expressed time period.

The implicit price for age represents a measure of depreciation. This coefficient would be expected to be negative and to remain fairly constant over time. However, as

the age variable in this data set is calculated as age in 1995 rather than age in the year of sale, the coefficient on age should become slightly more negative with each successive time period.

The coefficient on lot size represents the unit price of land. These are estimated directly for the naive models and expected to be positive. For the interactive models, these terms are not directly interpretable although, based upon predicted prices, an overall (urban) estimate of price per square foot of land for each time period can be derived. Unlike coefficient estimates for structural square footage and age, however, there are no obvious expectations as to how the unit price of land should behave over time.

It is unclear how the pattern of these coefficients over time will vary between each of the four models. However, the extent to which the temporal patterns of these coefficient estimates follow the above expectations will be the primary criterion for determining the superior model specification.

Consideration of Spatial Autocorrelation

A model specification that produces the best linear unbiased estimate assumes that error terms are not correlated; however, in the case of these cross-sectional price equations, the existence of spatial autocorrelation should be a distinct concern. Various methods are available for testing the spatial dependence of error terms. Here, a regression technique that is discussed by Cliff and Ord (1973) is employed:

$$\varepsilon_i = \theta + \rho \left[\Sigma_j^n w_{ij} \varepsilon_j \right] + \mu_i \tag{14}$$

where $w_{ij} = c_{ij} / \sum_j^n c_{ij}$ for connectivity matrix c_{ij} where i and j are adjoining (Queen's case join) areas. The constant, θ, is assumed to equal 0 and μ_i is a normal, random, and independent error term. Statistical tests which reject (the null hypothesis that) $\rho = 0$ indicate that correlation exists. Tests are performed for each time period in each model.

Spatial Variation in the Price of Land

Spatial variation in the price of land is designated only in the interactive models and is represented by a third order polynomial surface. While this may seem a rather rudimentary measure, the objective here is to capture a broad measure of the spatial variation in house prices over the urban landscape.

In an urban housing market, the demand for accessibility (to employment, shopping, schools, etc.) is extremely heterogeneous. This makes the evaluation of a land price surface enigmatic and constrains estimation to the simple third order surface that is employed here. However, this specification of land price should be sufficient to represent major spatial patterns in price and to observe changes in those patterns over time.

Interactive model coefficients represent the interaction of lot size with the various polynomial forms of (X,Y) coordinates. While these interactive coefficient estimates are not directly interpretable, they can be used to "predict" 3-dimensional land price surfaces.

Predicting Price and Appreciation

Approximate achievement of model expectations defined above along with diminished spatial dependence of the error terms will identify the superior model

specification and estimation method. That model specification is used to predict prices for each time period; predicted prices are then used to calculate appreciation rates. Appreciation is calculated as the average annualized change in price between (two-year) time periods and is therefore expressed as an average annualized rate.

Standard Housing Prices

House prices are predicted using a standard bundle of square footage, age, and lot size. These standard characteristics are simultaneously averaged over the urban area and over the different time periods and are shown in Table 4-1.

Table 4-1 Standardized Housing Characteristics

	MEAN
Structural Square Footage (SQFT)	1488
Age of Structure (AGE)	37.51
Square Footage of Land (LOT)	13,360

The interactive house price equation predicts house prices at different points in (X,Y) space; these price (trend) surfaces are demonstrated visually with 3-dimensional maps for each time period. Using the standard bundle, housing characteristics are held constant over space and, therefore, the house price surface at any point in time will replicate the land price surface.

Actual housing prices are likely to vary widely over the urban area. Because this model allows prices to be separated for land and structure, structural characteristics can

be held constant to observe the variation in price over time (i.e., appreciation) due primarily to location in space.

Temporal Implications

The superior model is then used to reveal the separation of appreciation into structural and locational elements. Land prices are averaged over space so that overall (temporal) structural appreciation can be compared to the overall temporal appreciation of land alone. The total "composite" (land plus structure) cumulative appreciation rate is then calculated; this (hedonic composite) index should approximate a cumulative appreciation rate derived from alternative methodologies such as an average price index or repeat-sales price index.

Finally, the model is used to predict (standard) house prices for the 140 points in space and, from those prices, infer average annualized rates of appreciation. A two-year appreciation rate is calculated between each time period; an average is then taken of those rates and annualized for all 140 points. Then, areas of predicted abnormal appreciation are identified.

This specification allows the observation of appreciation due solely to location, an approach that would not be possible with models that do not fully incorporate location. This may reveal appreciation characteristics that would be otherwise masked by the spatial or non-spatial variation in other housing attributes. The mixture of house size, age of structure, and lot size differs across the urban area and actual appreciation would be expected to be more erratic than predicted (constant quality) appreciation due to variation

in demand for non-locational attributes. Theory would suggest that age and lot size show more explicit spatial patterns while house size is more likely to be scattered and have less of a spatial pattern. While these characteristics may influence appreciation, this methodology predicts for the (constant quality) standard bundle and therefore measures the effects of "pure" spatial influences due only to location.

Patterns of Appreciation

Using the hedonic model with predicted appreciation as the dependent variable, both structural and spatial patterns of house price appreciation are investigated. First, structural characteristics are investigated. The work of deLeeuw and Struyk (1975) suggests that larger and newer houses will experience more rapid price appreciation; the (null) hypothesis that size and age do not influence appreciation will be tested with the following equation:

$$A_i = \beta_1 \, SQFT_i + \beta_2 \, AGE_i + \beta_3 \, LOT_i + \varepsilon_i \qquad (15)$$

where the average annualized appreciation rate A_i is expressed as a linear function of $SQFT_i$, the mean structural square footage, AGE_i, the mean age of the structure, and LOT_i, the mean lot size. Appreciation rates are regressed on these variables individually and in the multivariate equation above. The effect of house price (in 1980) is also investigated. As house price is assumed to be a linear function of square footage, age, and lot size, it is analyzed alone.

The existence of any abnormal appreciation may indicate that there is spatial variation in appreciation but it does not necessarily indicate any spatial pattern. Variation in appreciation may be explained not just by location but by demand for specific types of housing. However, with the predictive model, spatial patterns in appreciation may be more evident due to standard (constant quality across space) housing.

Regressing predicted appreciation on a polynomial expansion of (X,Y) coordinates will provide a test of the (null) hypothesis that no spatial pattern of house price appreciation exists. A third-order polynomial expansion of the TSA model, equation (8), is employed:

$$A_i = \Sigma_j^3 \Sigma_k^3 \, \beta_{jk} \, [X_i^j \, Y_i^k] + \varepsilon_i \qquad (16)$$

where A_i is the average annualized appreciation rate in grid cell i; β_{jk} denotes a vector of coefficients of X_{ij} and Y_{ik}, Cartesian coordinates of the grid cells and $j + k \leq 3$, where the model is a third order polynomial.

Although the coefficients in this equation lack any explanatory meaning, high statistical significance (of the coefficients) would indicate that spatial patterns do exist. Spatial patterns could be expected as A_i represents the average percentage difference between polynomial smoothed functions using a standard bundle of housing characteristics. The trend surface (TSA) equation is best represented visually; using computer graphics software, a 3-dimensional "appreciation" surface is created by graphing the equation.

Tests Using Repeat-Sales

Variation in appreciation due to location in space may be suggested by differences in the (interactive) hedonic model while spatial patterns may be implied by the TSA appreciation equation above. To verify the existence of any predicted abnormal appreciation, additional analyses are performed using the repeat-sales technique. These will test for any significant difference in price appreciation based upon individual houses which have sold twice.

The implicit assumption in the repeat-sales approach is that the quality of these houses has remained constant over time. Following Archer, Gatzlaff, and Ling (1995), the repeat-sales equation here estimates a dual index in an extension from equation (5) as follows:

$$\ln (P_{it} / P_{i\tau}) = \Sigma_t^T c_t D_{it} + \Sigma_t^T \hat{c}_t \check{D}_{it} + \varepsilon_{it} \tag{17}$$

where $\check{D}_{it}$ is a dummy variable which equals -1 at the time of initial sale or +1 at the time of second sale if the property is in an area of (predicted) abnormal appreciation, and 0 otherwise. Now, c_t is the logarithm of the cumulative price index in period t for the general market and $\hat{c}_t$ is the logarithm of any additional (positive or negative) cumulative appreciation due to being in an abnormal appreciation "submarket."

Areas of predicted abnormal appreciation may be indicated by the interactive model. Spatial patterns of appreciation may also be indicated by the TSA model,

prompting an analysis of the extent of abnormal appreciation. A spline technique is applied where multiple iterations of the model are run to estimate the distance effects of any abnormal appreciation.

The spline regression is a methodology which tests many (radial) distances to determine a "threshold" distance at which the difference between two areas is most pronounced. Here, distance intervals of 0.10 miles will be tested. The computer program for running the spline regression is included in Appendix D. The optimum model, based on coefficient t-statistics, will converge on a radial distance that contains a minimum number of observations and captures the greatest difference (in appreciation) between market and submarket. Repeat-sales tests based on individual sale transactions that verify the location of abnormal appreciation would strongly support the relevance of the interactive model. Indeed, the corroboration of model results at the individual house level with those from a generalized price model would have significant implications.

Methodology Summary and Assumptions

This methodology is based on the work of Jackson (1979) with substantial expansion. The methodology can be summarized in an 8-step procedure as follows:

1) Estimate the four model specifications, compare coefficient estimates, and identify the superior model specification to use for all prediction.

2) Predict and visually demonstrate land value surfaces.

3) Compare the appreciation of structural characteristics to the (spatially averaged) appreciation of land. Additionally, compare a composite price index (of land and structure) to other temporal price indices.

4) From the equations, calculate house price appreciation over space and identify areas of predicted abnormal appreciation.

5) Analyze appreciation as a function of housing (structural) characteristics as well as price (in 1980).

6) Analyze appreciation as a function of location, and visually demonstrate spatial patterns of appreciation.

7) Estimate the radial distances (about maximum and minimum points of predicted appreciation) at which houses within exhibit the greatest difference in appreciation from the rest of the market.

8) Test for statistically significant differences (between market and submarket) and visually graph a temporal price index for the market and any submarkets of abnormal appreciation.

Results for steps 1 and 2 are discussed in the following chapter. There, the superior model specification and estimation method is identified and used to predict land price surfaces. Results for the remaining steps are discussed in chapter 6.

This methodology has some limitations and also makes some explicit assumptions as to simplify the procedures and more easily interpret the results. Some basic definitions, limitations and assumptions are summarized as follows:

1) In this research the word "appreciation" can, as in the urban economics literature, refer to either appreciation (rising prices) or depreciation (falling prices). In Jacksonville during the 1980s, house prices were generally rising; however, the methodologies specified here can accommodate (and accordingly measure) both rising and falling prices.

2) The definition "abnormal appreciation" refers to prices that are rising at an appreciation rate that is above (positive abnormal) or below (negative abnormal) the average rate of appreciation. For the hedonic models, this is defined as 2 standard deviations from the mean, i.e., significant at the 0.05 level, assuming a normal distribution of appreciation rates. For the repeat-sales model, this is defined as statistically different from the market at the 0.05 significance level.

3) Prices are expressed in nominal dollars. Although prices are compounded forward (or discounted back) to the midpoint of the 24-month period using an urban house price index, there is no adjustment to real dollars. For the study of spatial variation in price appreciation, the use of real or nominal dollars is irrelevant.

4) The limitation of the study area to a 154-square-mile area has potential boundary problems in that major urban nodes or other important influences may be located just outside the study area. However, the polynomial expression of land price should reflect the influence of any external effects that are located outside the study area.

5) The structural variables of square footage and age are somewhat limited but they are the only structural variables available in the (Florida DOR) data set. However, as other studies have shown, these variables are the most important and are sufficient for the generation of hedonic indices (Gatzlaff and Ling, 1994).

6) The polynomial expression is rather limiting in its ability to estimate spatial variation in the price of land. Jackson (1979) employed a fourth-order model; preliminary tests of the data here suggest that only a third-order model will work well in all time periods. However, this expression should be sufficient to capture significant variation in house prices.

7) Many alternative functional forms of the estimating equation are available, including log-linear, semi-log, and Box-Cox transformation. However, preliminary tests suggest that such functional forms do not offer significant improvement over the linear/polynomial form that is specified here.

Additional definitions, limitations and assumptions are discussed elsewhere in the text where appropriate. For example, linear regression assumptions are discussed in Appendix B. Alternative solutions and suggestions are offered in chapter 7 under "Directions for Further Research." Results for the price equations are examined in the following chapter.

CHAPTER 5
PRICE EQUATION RESULTS

Price Model Comparison

Price equations are estimated for six different time periods using both single period estimation and simultaneous, seemingly unrelated, regression (SUR) estimation for both the naive and interactive models. Four model specifications are defined as follows:

1) naive, single-equation estimation (NSE)
2) naive, seemingly unrelated, regression estimation (NSUR)
3) interactive, single-equation estimation (ISE)
4) interactive, seemingly unrelated, regression estimation (ISUR)

These model specifications and estimation methods are compared with respect to coefficient estimates of structural (unit) prices and land (unit) prices, as well as spatially dependent error terms. The superior model specification will be used for prediction.

Model Specification

The aspatial or naive model is a standard, strictly cross-sectional hedonic price equation of the following form:

$$P_i = \beta_0 + \beta_1 \, SQFT_i + \beta_2 \, AGE_i + \beta_3 \, LOT_i + \varepsilon_i$$

Naive model variables are described as follows:

P_t the mean of actual transaction prices that have been compounded forward or discounted back to January 1 of the time period year at the overall urban rate of appreciation (each time period contains sales from two years)

$SQFT_t$ the mean structural square footage for the given time period

AGE_t the mean (1995) building age for the given time period

LOT_t the mean lot size for the given time period

This model specification is used in both the single-equation estimation (NSE) and simultaneous-equation estimation (NSUR) naive models. The spatial or interactive model interacts lot size with a polynomial expansion of (X, Y) coordinates as follows:

$$P_t = \beta_1\ SQFT_t + \beta_2\ AGE_t + \Sigma_j^k\ \beta_j \left[LOT_t * f_{tj}(X,Y) \right] + \varepsilon_t$$

Cartesian coordinates are the average X and Y coordinates. Because of multicollinearity problems, the interactive terms of L_XY_t, L_X3_t, and L_X3_t are dropped, leaving the following interactive terms:

L_X_t the product of lot size times X

L_Y_t the product of lot size times Y

L_X2_t the product of lot size times X-squared

L_Y2_t the product of lot size times Y-squared

L_X2Y_t the product of lot size times X-squared times Y

L_XY2_t the product of lot size times X times Y-squared

This (third order) model specification is used in both the single-equation (ISE) and simultaneous-estimation (ISUR) interactive models. Alternative functional (logarithmic) forms are tested for single-equation estimation models but offer no significant improvement. The linear model also provides more directly interpretable results.

Structural Unit Prices

Structural prices (per square foot) are assumed to be spatially constant in all models. However, estimated prices vary significantly between model specifications. The coefficients on square footage (in dollars) are shown in Table 5-1; these are all significant at the 0.001 level or better. Complete results are presented in Appendices E through H.

Table 5-1 Coefficients on Square Footage of Structure

	NSE	NSUR	ISE	ISUR
1980	25.16	21.98	24.08	23.57
1982	32.06	26.38	27.70	26.48
1984	38.99	30.32	31.18	30.02
1986	36.88	30.58	32.01	31.38
1988	44.32	36.33	33.38	32.48
1990	38.28	31.79	31.25	30.93

These coefficients can directly be interpreted as the price per square foot to construct a new house in the expressed time period. The simultaneous-equations estimation of the interactive model (ISUR) predicts a temporal index (based on estimated coefficients) that are more similar to the Producer Price Index (PPI) for construction

materials than the other model specifications. Square footage price coefficients for all model specifications (from Table 5-1) are converted to indices. All estimated coefficient prices are divided by the 1980 coefficient price; this generates cumulative indices that are set to value of 1 in 1980. In Table 5-2, these are compared to the PPI index which is adjusted (to value of 1 in 1980) in the same manner. The correlation coefficients between these model coefficients and the PPI index are then shown in Table 5-3.

Table 5-2 Indices for Square Footage and Producer Price Index

	NSE	NSUR	ISE	ISUR	PPI
1980	1.000	1.000	1.000	1.000	1.000
1982	1.274	1.200	1.150	1.123	1.095
1984	1.550	1.379	1.295	1.274	1.156
1986	1.466	1.391	1.329	1.331	1.184
1988	1.762	1.653	1.386	1.378	1.272
1990	1.521	1.446	1.298	1.312	1.346

Table 5-3 Correlation of Coefficients with Producer Price Index

	NSE	NSUR	ISE	ISUR
PP INDEX	0.835	0.875	0.850	0.881

The ISUR model specification predicts a temporal index (of coefficient prices on square footage) that most closely emulates the Producer Price Index for construction materials. These two indices demonstrate roughly the same cumulative appreciation (31 and 34 percent) in 1990. Additionally, the simultaneously estimated

interactive model is most highly correlated with the Producer Price Index for construction materials between 1980 and 1990.

Structural depreciation is estimated using the average age of houses; this variable is also assumed to be spatially constant. Again, estimation of coefficient prices varies between models as shown in Table 5-4; these coefficients represent dollars of depreciation for each additional year of house age and are all significant at the 0.001 level or better.

Table 5-4 Coefficients on Age of Structure

	NSE	NSUR	ISE	ISUR
1980	-220.92	-282.78	-217.91	-186.30
1982	-248.49	-344.11	-333.62	-266.91
1984	-285.00	-370.37	-385.11	-317.08
1986	-272.43	-351.93	-391.35	-335.39
1988	-112.36	-203.52	-396.08	-325.14
1990	-170.96	-219.37	-393.96	-317.11

The ISUR model specification predicts a temporal progression of coefficients that is more systematic than the other models. Specifically, this follows the expectation that, because the age variable in this data set is calculated as age in 1995 rather than age in the year of sale, the coefficient on age should become slightly more negative with each successive time period.

These coefficients can directly be interpreted as the amount of physical depreciation that occurred (on average) in the specific time period. To be expressed as a percentage, construction costs (square footage coefficients times average square footage)

are subtracted from house (structure only) prices predicted by the simultaneous-equations estimation of the interactive model. This yields a cumulative physical depreciation estimate of approximately 14 percent.

Land Unit Prices

Land prices (per square foot) are assumed to be spatially constant in the naive models but are allowed to vary spatially in the interactive models. Estimation of coefficient prices varies somewhat between model specifications and estimation methods; these are shown (in dollars per square foot) below in Table 5-5.

Table 5-5 Coefficients on Square Footage of Land

	NSE	NSUR	ISE*	ISUR*
1980	0.862	0.878	0.780	0.730
1982	1.163	1.192	1.232	1.151
1984	1.096	1.393	1.480	1.382
1986	1.531	1.671	1.804	1.690
1988	1.613	1.802	2.047	1.917
1990	2.059	2.226	2.382	2.174

* implied spatial average (not actual) coefficient

The coefficients shown above for the interactive models are calculated by predicting the land value at the (140) points in space and taking a spatial average. While these averages are in line with naive model results temporally, the focus of the investigation here is the variation of land values over space. Significant spatial variation

is found to exist; Table 5-6 demonstrates the variation (standard deviation and range) in land price (in dollars per square foot) over the (n = 140) grid cell space.

Table 5-6 Summary Statistics for Land Unit Prices over Space

	MEAN	STD. DEV.	MINIMUM	MAXIMUM
1980	0.730	0.121	0.163	1.032
1982	1.151	0.143	0.232	1.568
1984	1.382	0.130	0.712	1.799
1986	1.690	0.117	1.313	2.031
1988	1.917	0.130	1.516	2.148
1990	2.174	0.174	1.646	2.648

Spatial Autoregressive Variable

The interactive model will estimate an overall assessment of the intraurban variation in land price. To examine more "localized" effects, a spatial autoregressive price variable has been created; this variable is an average of price in all contiguous grid cells based upon the Cliff and Ord (1973) "Queen's case" (edge-to-edge and vertex-to-vertex) set of joins.

The spatial autoregressive variable is tested as an additional variable in the interactive models. Alone, this variable is statistically significant. However, when it is included as an additional variable along with square footage, age, and the interactive terms, it becomes insignificant in all years. This suggests that the interactive models, with their third-order polynomial expression of land prices, are sufficiently explaining the

spatial variation in house prices--or at least the spatial variation that can be estimated from the data available.

Spatial Dependence of Error Terms

Various methods are available for testing the spatial dependence of error terms; here, a regression technique is employed. Tests are performed for each time period in each model; t-statistics are presented in Table 5-7 where the critical value of t at the 0.05 significance level using a two-tail test is 1.98. These results demonstrate the ability of the ISUR model specification to reduce spatially autocorrelated error terms.

Table 5-7 t-statistics on Tests for Spatial Dependence

	NSE	NSUR	ISE	ISUR
1980	3.982	4.746	2.837	2.487
1982	2.562	4.979	1.978	0.751
1984	4.827	5.244	4.095	2.247
1986	5.885	7.116	4.477	2.625
1988	4.316	5.959	4.502	2.703
1990	5.430	6.061	3.172	1.977

Model Estimation and Prediction

The ISUR model specification is chosen as the best overall model and is used to predict house price variation over space. House prices are predicted for each time period by applying the estimated implicit prices to a standardized bundle of housing attributes.

Standardized housing characteristics of square footage, age, and lot size are calculated as the combined average over time and space for all houses.

Price Equations

The ISUR model specification produces simultaneous-equations estimation of structural and (interactive) locational model coefficients for all time periods. Structural coefficients are significant at the 0.001 level while interactive coefficients are nearly all significant at the 0.05 level; estimates are shown below in Table 5-8.

Table 5-8 ISUR Equation Coefficients

	1980	1982	1984	1986	1988	1990
SQFT	23.57	26.48	30.02	31.38	32.48	30.93
AGE	-186.30	-266.91	-317.08	-335.39	-325.14	-317.11
L_X	-168.63	-569.65	-499.74	-547.31	-398.33	-406.13
L_Y	478.1	1206.1	1056.9	1128.5	986.22	1031.1
L_X2	9.04	26.25	24.95	27.43	20.91	23.09
L_Y2	-43.55	-92.32	-80.87	-79.50	-70.24	-75.24
L_X2Y	-0.939	-2.194	-2.106	-2.215	-1.844	-2.070
L_XY2	1.913	3.955	3.626	3.549	3.044	3.335

Most obvious about the interactive coefficient estimates is their temporal pattern, that is, the coefficients demonstrate a non-random pattern over time. This strongly suggests that intraurban variation in the appreciation of urban land may be likely to have a spatial pattern.

Land Value Prediction

Intraurban variation in house price that is captured in the price equations can be demonstrated visually. The interactive coefficients can be multiplied by the appropriate (X,Y) coordinate expansion to predict a set of Z values that are associated with each (X,Y) point in space. Surface maps are created to demonstrate the spatial variation in urban land values. For the predictive (standard bundle) model, housing characteristics are held constant over time and space and therefore the house price surface would be identical to the land price surface; all Z values are simply shifted upwards by the value of a standard house.

The three dimensional land value surfaces are created using an 80 X 100 line grid to represent the 11 mile by 14 mile (154-square-mile) urban area. Therefore, the grid lines are spaced at approximately 0.14 miles. Input data for the construction of the surface maps consist of the 140 (X,Y) grid cell coordinates and their associated Z values, the predicted land values at those points. The surface maps are generated using a kreiging process which interpolates a smoothed set of Z values over space based on the uneven distribution of (X,Y) points.

While a series of land value surfaces may suggest a spatial pattern of appreciation, the equations for these surfaces will be combined to specifically calculate predicted appreciation rates over space in the following chapter. Specifically, appreciation will be calculated as the average annualized percentage difference between time periods. Areas of abnormal (greater or less than two standard deviations from the mean) will be depicted.

Additionally, these appreciation rates will be analyzed as a function of location and used to create an appreciation rate surface map.

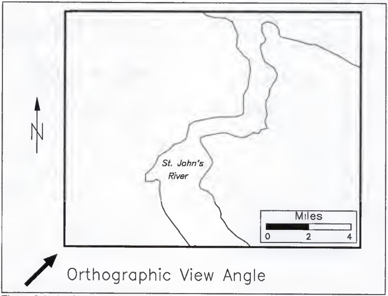

Figure 5-1 Angle of View for Surface Maps

Land value surface maps are created for all time periods, 1980 through 1990, and show the predicted price surface. The orthographic projection angle is shown above in Figure 5-1; this is a 225 degree rotation about the Z-axis with a tilt of 30 degrees. These maps view the study area from the southwest corner looking towards the northeast and are shown in Figures 5-2 through 5-7.

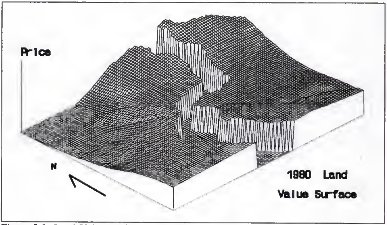

Figure 5-2 Land Value Surface for 1980

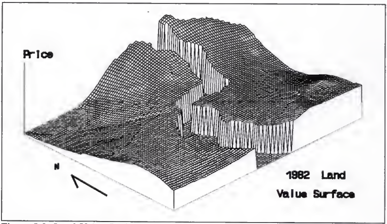

Figure 5-3 Land Value Surface for 1982

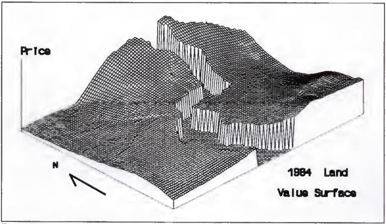

Figure 5-4 Land Value Surface for 1984

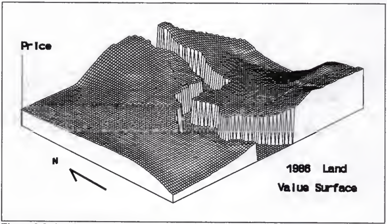

Figure 5-5 Land Value Surface for 1986

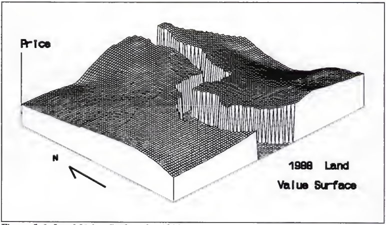

Figure 5-6 Land Value Surface for 1988

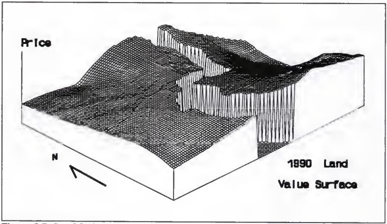

Figure 5-7 Land Value Surface for 1990

The third-order polynomial surfaces derived from the price equations and shown in Figures 5-2 to 5-7 capture only an abstract representation of the urban dynamics that occurred during this time span. The maps show the price surface--the spatial variation from the minimum to maximum value--and they would look identical for predicted land values only or for predicted values of land and housing. These land price surface maps demonstrate an obvious temporal pattern of above average price increase in the northwest and below average price increase in the northeast.

In the following chapter, the price equations are used to compare changes in the prices of structural characteristics to changes in the (spatially averaged) price of land. The price equations are then used to determine house price appreciation over space and identify areas of abnormal appreciation. Appreciation is next analyzed as a function of housing (structural) characteristics and location. Finally, the repeat-sales technique is used to verify the existence of abnormal appreciation and estimate the radial distances at which houses within exhibit the greatest difference in appreciation from the rest of the market.

CHAPTER 6
HOUSE PRICE APPRECIATION

Predicted Appreciation

The preceding chapter identifies the interactive model with simultaneous-equations estimation (ISUR) as the superior model specification with which to predict prices for each time period. In this chapter, prices are determined for land and structure separately so that the appreciation of each can be observed. The composite of locational and structural appreciation over time is then compared to alternative house price indices.

These predicted prices are then used to calculate appreciation rates where appreciation is calculated as the average change in price between time periods and is therefore expressed as an average annualized (two-year) rate. Areas of implied (positive or negative) abnormal appreciation are identified where abnormal appreciation is defined as appreciation above or below two standard deviations from the mean rate of appreciation.

Temporal Implications of the Price Model

Total prices for house, land, and their composite are predicted for a standardized urban house that has a living area of 1488 square feet, age of 37.5 years, and lot size of 13,360 square feet. To investigate temporal effects, predicted land prices are averaged

(over space) for each time period; this is accomplished by predicting land prices for all (X,Y) coordinates with the standard lot size and taking an average. Total prices of land, structure, and their composite are shown in Table 6-1. These prices, in dollars, for house (structural characteristics), land (location), and their composite are then all divided by their 1980 price and expressed as indices in Table 6-2.

Table 6-1 Estimated Total Prices

	HOUSE	AVERAGE LAND*	AVERAGE COMPOSITE*	STANDARD DEVIATION
1980	28922	10123	39045	12780
1982	30250	15712	45963	15377
1984	32727	18694	51421	16967
1986	33761	22471	56232	17711
1988	35770	25310	61080	18020
1990	33872	28889	62760	19845

* Average for urban area

Table 6-2 House Price Component Indices

	HOUSE	LAND	COMPOSITE
1980	1.000	1.000	1.000
1982	1.046	1.552	1.177
1984	1.132	1.847	1.317
1986	1.167	2.220	1.440
1988	1.237	2.500	1.564
1990	1.171	2.854	1.607

The composite index is based on predicted total prices and reveals the proportions of total price due to structure and land. For this time span, approximately 74 percent of total value is attributable to the structure while 26 percent is attributable to land; these are averages for the urban area. Additionally, intraurban price indices could be calculated based on the predicting equations for specific (X,Y) coordinates.

The composite price index predicted by the ISUR model specification, with land values averaged over the urban area, is equivalent to a standard cross-sectional hedonic index. This index is compared to an average house price index (based on all sales) and repeat-sales index (based on houses which sold twice) for the same (1980-1990) time period in the 154-square-mile urban area. These indices are shown in Table 6-3.

Table 6-3 Alternative House Price Indices

	AVERAGE PRICE	HEDONIC (ISUR)	REPEAT-SALES
1980	1.000	1.000	1.000
1982	1.166	1.177	1.150
1984	1.268	1.317	1.303
1986	1.414	1.440	1.462
1988	1.462	1.564	1.550
1990	1.528	1.607	1.601

The hedonic ISUR index is generated from aggregated biannual sales data. The average price index and repeat-sales index, however, are based on single year sales; only the alternate (even-numbered) years are shown above. Regardless, the hedonic index is similar to the average price index and nearly identical to the repeat-sales index.

The interactive model with simultaneous-equations estimation (ISUR) is shown to be a superior specification and methodology because coefficient estimates are more fitted to theoretical expectations. More importantly, the ISUR model specification convincingly produces a methodology for separating house price from land price and therefore allows the appreciation of those two components to be observed independently.

The hedonic ISUR (composite) index in Tables 6-2 and 6-3 indicates a cumulative appreciation rate of 60 percent, approximately 5 percent annualized; this is based on prices that are averaged over space. These predicted price indices suggest that structural appreciation (rise in cost of construction less physical depreciation) averaged about 1.6 percent annually while land appreciation averaged approximately 11 percent.

On a cross-sectional basis, the composite price appreciation average is 5.3 percent annualized with a standard deviation of about 0.5 percent. This is the average of appreciation rates for different points in space; cross-sectional variation in appreciation is the central focus of this investigation.

House Price Appreciation

The model is used to predict house prices for (the 140) points in space and, from those prices, to infer average annualized rates of appreciation. House price appreciation is predicted using a standard bundle of housing characteristics. These standardized characteristics are the average for the urban area over all time periods. The standard house has a living area of 1488 square feet, age of 37.5 years, and lot size of 13,360 square feet. Predicted appreciation is shown in Figure 6-1.

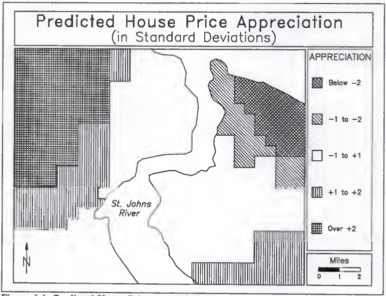

Figure 6-1 Predicted House Price Appreciation

Figure 6-1 shows predicted appreciation in standard deviations where appreciation is approximately normally distributed. The average appreciation rate is 5.3 percent with a standard deviation of 0.5 percent. Abnormal positive appreciation is defined as over 6.3 percent (2 standard deviations above the mean) while abnormal negative appreciation is defined as under 4.3 percent (2 standard deviations below the mean). An area of predicted abnormal positive appreciation is apparent in the northwest corner of the study area while an area of abnormal negative appreciation is evident in the northeast.

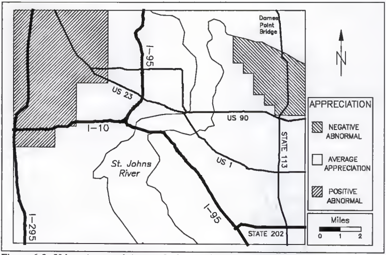

Figure 6-2 Urban Axes and Areas of Abnormal Appreciation

Figure 6-2 shows the areas of predicted abnormal appreciation in the northwest (positive) and the northeast (negative) along with major urban axes. Urban axes increase accessibility to certain areas--in the 1980s, areas in the proximity of I-295, in the northwest, and the Dames Point Bridge, in the northeast, realized a change in accessibility benefits from the construction of these axes. Households in the northwest receive the benefit of increased accessibility to the Jacksonville International Airport, other interstates, and the urban center (via I-10). However, households in the northeast receive a negligible or even negative benefit due to the Dames Point Bridge. Changes in accessibility benefits there (access to the urban center) are realized more by households on the north side of

the St. John's River; the south side may have experienced only more congestion and perhaps more crime, an accessibility dis-benefit.

However, as other work has strongly suggested, it is not highway access alone that increases the demand for individual sites. For example, in Brigham's (1965) investigation, an accessibility potential (to employment) variable is considered in addition to distance gradients alone. The locations of those places being accessed (i.e., employment centers, schools, shopping centers, etc.) are, therefore, important in such house price models.

This model specification avoids the need to know the locations of important nodes; by specifying the price of land as a polynomial expression, all external effects are implicitly included in the price of housing. Housing characteristics, however, may have an effect on appreciation. Averages of housing characteristics are shown in Table 6-4.

Table 6-4 Housing Characteristic Averages for Areas of Abnormal Appreciation

	NORTHWEST	MARKET	NORTHEAST
1980 PRICE	28435	39400	48230
SQUARE FOOTAGE	1220	1488	1705
AGE	37.98	37.49	29.97
LOT SIZE	8612	13360	15462

There are significant differences between housing characteristics in the northwest and northeast and the overall market. In the following section, housing characteristics are analyzed as potential explanatory factors of house price appreciation.

Appreciation Equation Results

A hedonic model with appreciation as the dependent variable is used to investigate both structural and spatial patterns of house price appreciation. Variation in appreciation may be attributable to housing characteristics of square footage, age, and lot size, location in space, or their combination. Because appreciation is calculated as the average percentage change in predicted prices using standard characteristics and a polynomial expression of land price, spatial patterns may be anticipated.

Characteristic Effects

The effect of structural and lot size characteristics is investigated by regressing appreciation on actual characteristics for the 140 points in space. The variables are analyzed individually and together as follows:

$$A_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \text{and} \qquad A_i = \beta_0 + \Sigma_j^k \beta_j X_{ji} + \varepsilon_i$$

where β_j, 1 to k, represent the characteristic prices of square footage, age and lot size.

Square footage and lot size (in thousands of square feet) are found to have had a negative influence on appreciation while age appears to have had a positive influence. These effects are observed in both a univariate and multivariate regression equations. In the multivariate equation, square footage becomes statistically insignificant, due most likely to its relatively high correlation ($r = 0.85$) with lot size.

Coefficient estimates for all characteristics are extremely small; large unit changes would have negligible effects on appreciation. Additionally, R-squared statistics on these equations are small, ranging between 0.10 and 0.20. So, while these characteristics may have a small significant effect, they explain very little of the variation in appreciation. Table 6-5 shows basic results (coefficients and associated t-statistics) of the individual (univariate) and joint (multivariate) regressions; full results are given in Appendix I.

Table 6-5 Effects of Structural and Lot Size Characteristics on Appreciation

	UNIVAR. COEFF.	UNIVAR. T-STAT.	MULTIVAR. COEFF.	MULTIVAR. T-STAT.
SQFT	-.0000078	-4.756	.0000005	0.170
AGE	.0002407	4.143	.0001476	2.517
LOT	-.0006627	-5.699	-.0005848	-2.616

These statistics suggest that smaller and older houses experience greater price appreciation; a finding which contradicts the work of deLeeuw and Struyk (1975). However, these results, along with results from the price model strongly suggest that it is the demand for accessibility (location) rather than the demand for specific types of housing that is the primary driver of intraurban house price appreciation.

Effects of Price

The above results suggest that, while the relationships between house price appreciation and housing characteristics are statistically significant, estimated coefficients

(interpretable as the increase in appreciation for a one unit change in the characteristic) are very small and have a negligible effect. Housing markets may be segmented by size or age, but also by price. Indeed, house price has been shown to be a linear function of house size, age, and lot size. As a basis for market segmentation, price is intuitively appealing because it is the basis for demand. Households with different income levels are limited, at least on the upper end, to the range of house prices they can afford.

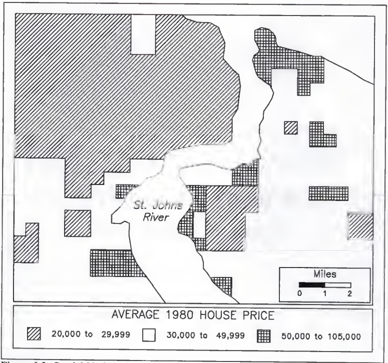

Figure 6-3 Spatial Variation in House Price

The spatial distribution of (1980) house price is shown in Figure 6-3. Most noticeably, the more expensive houses tend to be located along the eastern and southwestern shores of the St. John's River; these houses are located in Jacksonville's more exclusive residential neighborhoods. They also tend to be larger; house price is highly correlated ($r = 0.935$) with house size. To test the effects of price on appreciation, appreciation rates are regressed on (1980) house prices for the 140 grid cells. Table 6-6 shows basic statistical results.

Table 6-6 Statistics for the Regression of Appreciation on Price

	COEFFICIENT	STD. ERR.	T-STATISTIC
PRICE80	-.00000019	.000000036	-5.360

Price is significant but its effect is small; the coefficient above suggests that for a thousand dollar increase in price, appreciation will decrease by only 0.02 percent. Such market segmentation should not be ignored however. Variation in the demand (over different price levels) for housing may influence appreciation and cloud studies seeking to determine if such variation is explainable.

However, to investigate spatial patterns of appreciation, housing characteristics such as price (or size) are not considered; appreciation rates are based on the predictive (standard bundle) model where all houses are assumed identical. Thus, spatial patterns of land appreciation--if they exist--will be more obvious and not masked by variation in housing attributes; this is discussed next.

Spatial Patterns of Appreciation

To investigate the variation in appreciation due to location in space, appreciation rates are regressed on a polynomial expansion of (X,Y) coordinates. A stepwise procedure (see Appendix B) selects the most significant variables and drops those which are likely to cause multicollinearity. The best fitting (TSA) equation is as follows:

$$A_i = \beta_0 + \beta_1 X_i + \beta_2 Y_i + \beta_3 X_i^2 + \beta_4 X^2Y_i + \beta_5 XY_i^2 + \varepsilon_i$$

Although the TSA equation lacks any explanatory meaning with regard to direct interpretation of the coefficients, high statistical significance would indicate that spatial patterns do exist. Table 6-7 shows t-statistics; full results are given in Appendix I.

Table 6-7 t-statistics for the TSA Appreciation Model

	X	Y	X2	X2Y	XY2
T-STATISTIC	-18.598	61.225	40.572	-70.692	7.677

The high statistical significance of the coefficients indicate that spatial patterns indeed exist; the critical value of t at the 0.05 (two-tail) significance level is 1.98. The adjusted R-squared statistic is 0.996, indicating that the overall explanatory power of the model is exceptional. However, TSA equations are best demonstrated visually; the estimated equation is graphed in Figure 6-4.

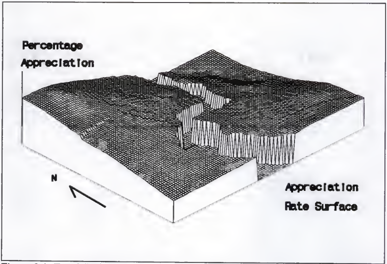

Figure 6-4 Trend Surface Analysis of Appreciation

The appreciation equation exhibits a spatial pattern that agrees with prior observation; abnormal positive appreciation is evident in the northwest corner of the study area while an area of abnormal negative appreciation is obvious in the northeast. Additionally, significant appreciation can be observed in the southeast, although this has not been identified as "abnormal." Most importantly, however, is the manifestation of a very distinct spatial pattern of appreciation.

This manner of calculation makes suggested abnormal appreciation suspect; predicted appreciation is the average of differences between smoothed polynomial functions that themselves are based on averages of actual house prices. However, repeat-sales at the individual house level can be used to test these indications.

Repeat-Sales Results

To test for the existence of the predicted patterns of positive and negative abnormal appreciation, additional analyses are performed using the repeat-sales technique. These tests will identify any significant difference (between the urban market area and a locational submarket area) in price appreciation based upon individual houses which have sold twice. The estimating equation is as follows:

$$\ln (P_{it} / P_{i\tau}) = \Sigma_t^T c_t D_{it} + \Sigma_t^T \hat{c}_t \check{D}_{it} + \varepsilon_{it}$$

where the coefficient estimate $\hat{c}_t$ is the logarithm of any additional (positive or negative) cumulative appreciation due to being in a specific submarket.

Of the 11,570 sales which were aggregated over both space and time, there are 3,998 houses which sold twice. Not only is the data set significantly different, but the repeat-sales technique is an entirely different methodology; the model provides the advantage of using full information of the individual observation, thus enabling the observation of locational effects on appreciation at the individual house level.

Verification of abnormal appreciation using the repeat-sales technique would strongly support the validity of the simultaneously estimated interactive hedonic model and the TSA appreciation model. Where the interactive model predicts appreciation from the "differences" in generalized price surfaces, the repeat-sales model is based upon the appreciation of houses in the overall market and specific submarket areas.

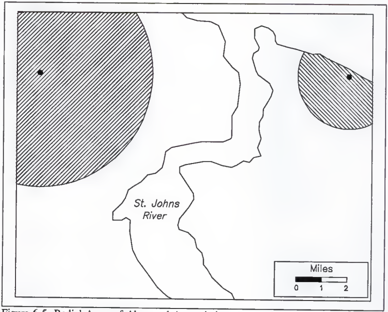

Figure 6-5 Radial Areas of Abnormal Appreciation

Because spatial patterns of appreciation are evident in the TSA model, a spline regression procedure is employed to estimate the distance effects of any abnormal appreciation. In the spline regression, multiple iterations of the model are run to determine the radial distance (about the maximum and minimum appreciation values) at which appreciation is most different between market and submarket. For the northwest (predicted positive abnormal), the radius is 4.4 miles while for the northeast (predicted negative abnormal), the radius is 2.0 miles. These radial areas identified by the spline regression are shown in Figure 6-5, above.

The repeat-sales model is used as a test of differences between the specific submarket and the overall market. Submarket appreciation is considered "abnormal" if a statistically significant difference exists between the two. The standard statistical hypothesis that each $\hat{c}_t$ equals zero is evaluated. However, the (null) hypothesis must be rejected for several years to assume any pattern of abnormal appreciation. More importantly, the pattern of cumulative appreciation differences should increase (decrease) over time to support the perception of aberrant appreciation.

The 4.4 mile radial area in the northwest is constrained by the study area boundary. This is a 31.3 square mile area that contains 185 observed repeat-sales. Table 6-8 demonstrates that appreciation in the northeast is significantly more than the rest of the market; the critical value of t at the (one-tail) 0.05 significance level is 1.64.

Table 6-8 Submarket Appreciation in the Northwest

	COEFF.	T-STAT.	INDEX	MARKET	DIFF.
1980	------	------	1.000	1.000	0.000
1981	0.028	0.758	1.178	1.089	0.031
1982	0.028	0.620	1.178	1.151	0.027
1983	0.065	[illegible]	1.303	1.221	0.027
1984	0.047	1.353	1.353	1.305	0.063
1985	0.070	1.995	1.471	1.471	0.100
1986	0.097	2.936	1.602	1.451	0.148
1987	0.065	[illegible]	1.615	1.512	[illegible]
1988	0.104	3.019	1.718	1.548	0.170
1989	0.072	2.042	1.699	1.581	0.118
1990	0.050	1.439	1.694	1.611	0.082

In the northwest, abnormal positive appreciation is evident in all years with an average annual difference of 9.2 percent. It is statistically significant in 6 years at the 0.05 significance level using a one-tail test. The pattern of annual differences in appreciation is rather erratic; annual appreciation can be inferred from the difference column in Table 6-8. This index, along with the indices for the market and the northeast are graphically illustrated in Figure 6-6. As can be observed there and above, the cumulative difference follows a steadily increasing pattern between 1982 and 1988.

The 2.0 mile radial area in the northeast is constrained by the St. Johns River and the study area boundary. This is a 6.4 square mile area that contains 470 observed repeat-sales. Table 6-9 shows that appreciation in the northeast is significantly less than the rest of the market; again, the critical value of t at the 0.05 significance level is 1.64.

Table 6-9 Submarket Appreciation in the Northeast

	COEFF.	T-STAT.	INDEX	MARKET	DIFF.
1983	-------	-------	1.000	1.000	0.000
1983	-0.024	-0.734	1.065	1.091	-0.026
1982	-0.033	-1.104	1.118	1.155	-0.037
1983	-0.004	-0.142	1.221	1.226	-0.005
1983	-0.032	-1.092	1.22[illegible]	1.311	-0.041
1985	-0.062	-2.103	1.299	1.383	-0.084
1985	-0.066	-2.382	1.377	1.377	-0.094
1987	-0.091	-3.105	1.395	1.527	-0.132
1988	-0.101	-3.287	1.421	1.571	-0.150
1989	-0.107	-3.615	1.439	1.602	-0.163
1990	-0.096	-3.034	1.477	1.626	-0.148

In the northeast, abnormal negative appreciation is apparent in all years with an average annual difference of -8.8 percent. It is statistically significant at the 0.05 level in six years. As demonstrated in Table 6-9 above and illustrated in Figure 6-6 below, the cumulative difference grew consistently larger between 1984 and 1989.

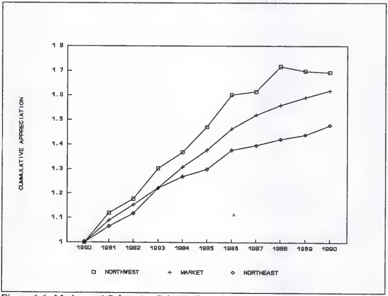

Figure 6-6 Market and Submarket Price Indices

The graphed indices in Figure 6-6 display a strong negative deviance from the market in the northeast and a less consistent, but statistically significant, positive deviance from the market in the northwest. Complete regression results for the northwest and

northeast areas are presented in Appendix J. These results strongly support the conclusions from the predictions of price and appreciation (from the hedonic, simultaneously estimated, interactive model).

The areas of predicted abnormal appreciation (based on standard deviations from the mean rate of appreciation) are similar to the radial distances estimated by the spline regression. In part, the "optimal" radial distance is a function of the number of houses contained within it; the submarket must contain a minimum number of observations (at least 100 of the 3998 total observations) but not so many as to lessen differences with the rest of the market. Thus, the sparsely developed northwest is designated a larger (4.4 mile) radius while the more densely developed northeast has a smaller (2.0 mile) radius.

Abnormal appreciation has been identified by a generalized model using a rudimentary third-order polynomial expression of land price and aggregate data. The validity of the implications (with regards to spatial variation in house price appreciation) of this model is confirmed by an alternative model (repeat-sales) that specifies appreciation as a function of time alone and uses individual sales observations. A conclusion and summary of these findings follows in the succeeding chapter.

CHAPTER 7
CONCLUSION

House Price and Appreciation

From the Brigham (1965) macro-analysis of Los Angeles County (4120 square miles) to the Johnson and Ragas (1987) micro-analysis of a 1.38 square mile area in New Orleans, empirical investigation has found strong support for the hypothesis that location (or accessibility) advantages are captured in the land price. However, house price models discussed in the literature today typically underspecifiy the characteristics of location. In the hedonic or repeat-sales equation, areal differentiation using dummy variables can be used to specify areal units such as census tracts; this is only a discrete measure of location that disregards potential boundary problems and provides no measure of any spatial pattern. Hedonic models may include distance gradients to capture distance decay effects (a spatial pattern) but, like the areal differentiation approach, can capture only limited aspects of location.

Even multinodal models cannot fully define the properties of location as an indeterminable number of ever changing externalities exist on the urban landscape. Such models are appropriate for analyses of particular locational effects but fail to capture the aggregate effect of location on the price of housing. Location as a "service bundle" is distinct; each site is unique with respect to its access to the urban environment.

The simultaneous-equations estimation of the interactive model provides a methodology to fully capture the effect of spatial variation in a hedonic model, at least on a broad level. While the third-order polynomial specification limits estimation to a very generalized land value surface, it is sufficient for the purposes of this study. Indeed, demand for accessibility in the urban housing market is so heterogeneous that the third-order specification is the only functional form that works well over many time periods and for that reason is selected here. For other land uses, such as office or retail, demand for accessibility may be expected to be more consistent, and thus, the use of higher order functions more appropriate.

Another important observation of the ISUR model specification is that house price information can be regarded not only as cross-sectional but as time-series as well. The simultaneous-equations estimation of house price equations using the SUR procedure is shown to provide a significant gain in the efficiency of the hedonic model. This model specification produces more reliable coefficient prices (of square footage and age) and reduces both contemporaneous and spatial autocorrelation of the error terms.

The ability of this model to predict intraurban house prices may also prove beneficial for other uses such as the setting of mortgage loan limits. The simultaneously estimated interactive hedonic model may provide a methodology more reliable than those (median and constant quality) discussed by Hendershott and Thibodeau (1990); the model has definite advantages with respect to the generation of submarket indices.

However, the major contribution of this model is its ability to reveal the separation of appreciation into structural and locational components. The fundamental deduction

here is that structural appreciation is due predominately to the rising cost of building materials (less physical depreciation), thus implying that the majority of all differences in real appreciation are due to changes in the relative values of location. This perception agrees with the classical economic ideology that property values are the residual effect of land.

By holding structural characteristics spatially constant and using a standard bundle of characteristics, the model allows the observation of appreciation due solely to location. While implicit characteristic prices on square footage and age vary over time, they remain consistent spatially. It is important to realize that such an investigation could not be properly conducted with models that do not fully incorporate location in this manner.

This approach also reveals appreciation characteristics that may otherwise be masked by the variation in other housing attributes; with the predictive model, spatial patterns in appreciation are more evident. Variation in appreciation may be explained not just by location but also by demand for specific types and prices of housing. House price, house size, the age of the structure, and other housing characteristics vary across the urban area and actual appreciation may be expected to be more erratic than the predicted (constant quality) appreciation due to variation in demand for non-locational attributes. Thus, methodologies that do not incorporate measures of location in such a manner may lack the ability to uncover various idiosyncrasies of house price appreciation.

The fundamental deduction of this investigation is that house price appreciation varies over the urban area in a spatially and temporally consistent manner. Such variation is due to the underlying aggregate demand for accessibility benefits; these benefits change

in spatial patterns over time. In an urban housing market, the concept of accessibility is quite ambiguous; it differs for different households. The model specification used here captures only an abstract representation of the urban dynamics that occurred during this time span. This (third-order) function can identify only one absolute maximum and one absolute minimum point of appreciation; in actuality, many relative maximas and minimas may be expected.

This research finds no support for theoretical (ex-ante) appreciation expectations with regards to perimeter location, house age, or house size; to the contrary, it is found that smaller (and older) houses tended to appreciate more, but only negligibly so. Rather, it is found that house price appreciation is primarily affected by location and the changes in accessibility benefits at different locations; these benefits are priced by the market for residential housing. The changes in accessibility benefits are likely due to new urban nodes and axes or the changing influences of existing nodes and axes. The findings here support price appreciation implications from price theory: accessibility benefits are capitalized in the price and therefore, relative changes in accessibility benefits affect the level of change (appreciation) in price.

The primary contribution of this analysis is a methodology which reveals the appreciation maximum and minimum and determines the distance from those points at which appreciation (for the submarket within that radial area) is most different from appreciation for the rest of the market. While this captures the spatial pattern of appreciation and identifies abnormal appreciation in a very general manner, findings are substantiated by the evidence from repeat-sales.

The repeat-sales model provides strong support for the simultaneously estimated interactive model and its ability to predict areas of abnormal appreciation. The simultaneously estimated interactive hedonic model aggregates data both spatially and temporally and is smoothed over space by its polynomial functional form. The repeat-sales methodology, on the other hand, preserves full information of the individual house sale by combining sales data over different holding periods to estimate an annual index or sets of intraurban indices.

Appreciation in Jacksonville

This research has found that definite spatial pattern of house appreciation were apparent in Jacksonville during the 1980s. Abnormal positive appreciation was estimated in the northwest corner of the study area. This above-average appreciation is most easily explained by the urban axis, I-295, which was completed in the early 1980s and increased accessibility in the northwest. Households in the northwest received the accessibility benefits of increased access to the Jacksonville International Airport and other interstates, as well as improved access to the urban center.

Abnormal negative appreciation was estimated in the northeast corner of the study area. This below-average appreciation is most easily explained by another urban axis, the Dames Point Bridge, but for different reasons. Although construction was not completed until the late 1980s, commercial activity increased in the northeast corner of the study area in anticipation of the new bridge. This commercial activity and the additional congestion that it brought to the area was most likely a principal reason for lower house

price appreciation in the northeast. In fact, the increased access from the north side of the St. John's river and the increased commercial activity have been major factors in making this area the highest crime district in Jacksonville.

The price surfaces in Figures 5-2 through 5-7 and the appreciation surface in Figure 6-4 also identify an emerging urban node in the southeast corner of the study area although this area was not identified as having "abnormal" appreciation. During the 1980s, the Southpoint Business Park and Mayo Clinic were constructed; as employment nodes, these appear (at least visually) to have had an impact on housing prices.

For Jacksonville, housing characteristics were statistically significant factors or house price appreciation, as was house price (in 1980). However, coefficients were very small and these factors appear to have had a negligible effect. It is quite conceivable that there was a greater demand for lower-priced housing in Jacksonville in the 1980s and it is unknown what other factors may have influence housing prices in Jacksonville.

The conclusion from this investigation of house price appreciation in Jacksonville is that housing characteristics, including price, have little effect on appreciation. Rather, it is the changes in accessibility benefits that appear to be the fundamental cause. The works of Brigham (1965), Jackson (1979), Johnson and Ragas (1987), and others have strongly suggested that land prices are a function of accessibility. This research suggests that spatial variation in house price appreciation is essentially due to changes in accessibility, the result of the changing influence of nodes and axes that are integral parts of the ever-changing urban spatial structure.

Directions for Further Research

Much room remains for improvements on and extensions of the methods used here. The size of the study area is a primary interest, especially regarding the application of polynomial expressions of land price. Smaller areas may accommodate higher-order polynomial functions as evidenced by the Johnson and Ragas (1987) sixth-order function that was applied to a 1.38 square mile area. The defined (154-square-mile) study area that is used here is not expected to pose specific boundary problems as the polynomial expression should theoretically capture the external effects of any influences that are inside or outside the study area. However, it may prove interesting to investigate the application of different (third-order and higher) polynomial functions to a larger areas such as the entire county.

Spatial aggregation is another area of interest. Statistically, more observations (and thus more degrees of freedom) are desirable and will produce stronger results. The 140-grid-cell aggregation technique that was used here produced better results than aggregation at the census tract or census block group level. The notion of a spatial moving average is also intriguing; such an approach was justified by Brigham (1965) as a way to remove as much spurious variation (in house price) as possible and allow the investigation of general (rather than local) variations in land values. Brigham's moving average was one-dimensional (along a vector), but a two-dimensional moving average could be applied utilizing GIS. This technique was rejected here because of the double counting of some sales. However, to the extent that such double counting is random, this

technique could be justified and would result in a larger number of observations. In any aggregation technique, the number of individual houses being aggregated is also a concern and may influence results.

The structural characteristics of square footage and age were held spatially constant here with reasonable justification. However, the built form of housing may have spatial effects; these could be investigated by specifying square footage as spatially variant. Alternative functional forms could also be further investigated.

This research provides a rudimentary methodology for continuing investigations of intraurban variation in house price appreciation. The existence of abnormal appreciation however, does not imply overall abnormal returns. The total return on an asset is its appreciation (or capital gain) plus its rent (or dividend) yield. Thus, the relationship of house price appreciation with implicit house rents is an area for investigation. Additionally, the relationship with various measures of risk (including variance of appreciation and number of sales) remains an interesting research area.

This research fills a niche in the housing literature. The principal contribution is a methodology for investigating house price appreciation in a manner that fully incorporates location and separates the value due to location from the value due to structural characteristics. As suggested above, there is much room for further addition to and expansion of this work. Expanding on implications from the house price literature, this research also provides support for theoretical axioms of spatial variation in house price appreciation.

APPENDIX A
DATA PROCEDURES

The data for this study will come from the Florida Department of Revenue's (DOR) multi-tape database of county property tax records.

Procedure 1: Export DOR data

Read in data tapes to files extracting the following data in DOS (ASCII) format:

Field No.	Field Label
01	Parcel ID
04	D.O.R. Land Use Code
06	Total just value
07	Total assessed value
10	Land value
11	Land units code
12	Number of land units
15	Year improvement built
16	Total living area
21	Most recent sale price
22	Most recent sale date
28	Previous sale price
29	Previous sale date
42	Homestead exemption
51	Address1
52	Address2
53	City
54	State
55	Zip code

Procedure 2: Construct initial data set

Step 1: Identify those properties which are improved single family detached units (D.O.R. Code = 1), and delete all others.

Step 2: Identify those properties which have sold at least once in the 10 year study period (1979 - 1990), and delete all others.

Step 3: Identify those properties which are owner occupied (by homestead exemption) and delete all others.

Step 4: Clean data set of apparent error, abnormal, and incomplete observations:

A) Delete if land value < 1000.
B) Delete if assessed value < 1000.
C) Delete if year built < 1900.
D) Delete if total living area < 800.
E) Delete if total living area > 6000.
F) Delete if either sale price < 10,000
F) Delete if either sale price > 500,000

Step 5: Generate price per square foot variable for each year and calculate mean and standard deviation. Delete all observations greater than or less than 2.5 standard deviations from their respective means.

Procedure 3: Clean address data for geo-coding

Step 1: For ease of data manipulation in geo-coding it is easiest to split the dataset into two files; the first containing property specifications, prices, and dates (fields 04 to 29), and the second containing address information (fields 51 to 55), with both using Parcel ID (field 01) as the key field. Once geo-coded, the coordinate file can be merged with the initial data set.

Step 2: Arrange all data in following format:

NAME:	Parcel ID
ADDRESS:	Combine Address1 and Address2 so that this field contains a house number, directional prefix or suffix if applicable, street name, and street type suffix. For more accurate matching, this must be as complete as possible.
CITY:	City or town
STATE:	Two letter state abbreviation
ZIP:	5 digit zip code (5 characters only)
PLUS4:	4 additional digits (if available)

Step 3: Using a (GIS) address-matching procedure, determine latitude / longitude coordinates.

Step 4: Based on the an origin at the southwest corner of the county, convert latitude / longitude coordinates to Cartesian coordinates.

Procedure 4: Construct time period data sets

Step 1: Create separate data sets for biannual time periods (1979 and 1980, 1981 and 1982, etc.)

Step 2: Compound first year sales forward and discount second year sales at urban appreciation rate based on month of sale. That is, all 1979 sales are compounded and all 1980 sales are discounted to January 1, 1980.

Step 3: Using GIS, determine areal units which contain minimum number of sales for each time period.

Step 4: Using a (GIS) point in polygon procedure, aggregate individual property characteristics into areal units.

Procedure 5: Construct Repeat-Sales data set

Step 1: Identify those properties which have sold at least twice in the 10 year study period (1979 - 1990), and delete all others.

Step 2: Calculate time between sales (based on year and month of sale) and delete all observations with a holding period of under one year.

Step 3: Calculate annualized appreciation for each property. Then calculate mean and standard deviation for annual appreciation and delete all observations greater than or less than 2.5 standard deviations from the mean.

Step 4: Generate dummy variables for time of sale where $D_t = 1$ if t is most recent sale year, -1 if t is previous sale year, and 0 otherwise, for all (t) time periods.

Note: During specific repeat-sales analyses, distances from predicted appreciation maximum and minimum points to all houses will be calculated. Dummy variables for location within a radial area of (predicted) abnormal appreciation (where $\check{D}_i = 1$ if property is located in a certain area, and 0 otherwise) will also be created.

APPENDIX B
REGRESSION ASSUMPTIONS

Ordinary least squares (OLS) regression equations are used extensively in this research. Multiple linear regression is a powerful tool that has some implicit assumptions and potential problems; the econometric solutions to these problems can be quite complicated. Basic (OLS) assumptions, diagnostics, and remedial measures are as follows:

(1) Multicollinearity:

It is assumed that there is sufficient variation among the explanatory variables. A symptom of multicollinearity is a high R-squared but low individual t-statistics. This assumption can be checked by analyzing a correlation matrix of the explanatory variables. If two variables are highly correlated, it is quite possible that one is a function of the other (such as LOT_X and LOT_X2), and that one must be eliminated.

Often the most efficient solution is to use a stepwise regression. This enters variables into the equation based on a critical F-value and allows those variables to stay in the equation only if they maintain a critical F-value as other (significant) variables are entered into the equation. The stepwise procedure is employed during the investigation of spatial patterns of appreciation where appreciation rates are regressed on a polynomial expansion of (X,Y) coordinates.

(2) Linearity of functional form:

It is assumed that the relationship between dependent and independent variables is linear. It is quite possible that the relationship between price and square footage could be nonlinear; as square footage increases so does price, but at a decreasing rate. This is best analyzed by plotting the variables in question. Here, the relationship between price and square footage (and age) is assumed to be linear while the relationship between price and location is specified as a polynomial function.

(3) Heteroscedasticity:

It is assumed that the regression variance is constant. However, with cross-sectional data, it is not unusual to find an increasing variance. As price increases over space, its variance might also increase. A plot of the residuals provides the easiest verification; alternatively, the Goldfeld-Quant, Park, or Glejser tests can be performed. In cross sectional studies, the appearance of homoscedastic errors may actually be spatially autocorrelated errors.

(4) Serial Autocorrelation:

It is assumed that the regression error terms are random (not correlated). However, this can be a problem in time-series regression because variables are often correlated to some degree with themselves over time. This is commonly checked by testing the Durbin-Watson statistic.

In this research, a major insight is that the (cross-sectional) data is time-series as well. Individual equations are cross-sectional but these equations are run over multiple time periods. Contemporaneous correlation (correlation at a given point in time but not necessarily over time) in seemingly unrelated regression equations and a corrective procedure are discussed in Appendix C.

(5) Spatial Autocorrelation:

Again, it is assumed that the regression error terms are random (not correlated). This can be a problem in cross-sectional regression because variables are often correlated to some degree with themselves over space. This can be tested with a variety of means including Moran's I and Dacey's contiguity test. Here, a regression technique that is discussed by Cliff and Ord (1973) is employed; this is shown on page 39. Spatial autocorrelation is often the result of model misspecification; the interactive model that is employed here is shown (see Table 5-7 on page 55) as a method of reducing spatially autocorrelated error terms.

APPENDIX C
SEEMINGLY UNRELATED REGRESSION EQUATIONS

The seemingly unrelated regression equations (SUR) procedure simultaneously estimates a set of equations in a manner that may reduce contemporaneous (at a given point in time but not necessarily over time) correlation and improve the efficiency of the model. Consider a simplified version of the strictly cross-sectional house price model defined in equation 1 on page 17:

$$PRICE_t = \beta_j X_{jt} + \varepsilon_t$$

This model can be run separately for the six time periods (1980, 1982, 1984, 1986, 1988, 1990) in single estimation. Alternatively, the equations can be stacked and estimated simultaneously:

$$\begin{vmatrix} PRICE80 \\ PRICE82 \\ PRICE84 \\ PRICE86 \\ PRICE88 \\ PRICE90 \end{vmatrix} = \begin{vmatrix} X80 & X82 & X84 & X86 & X88 & X90 \\ X80 & X82 & X84 & X86 & X88 & X90 \\ X80 & X82 & X84 & X86 & X88 & X90 \\ X80 & X82 & X84 & X86 & X88 & X90 \\ X80 & X82 & X84 & X86 & X88 & X90 \\ X80 & X82 & X84 & X86 & X88 & X90 \end{vmatrix} \begin{vmatrix} \beta 80 \\ \beta 82 \\ \beta 84 \\ \beta 86 \\ \beta 88 \\ \beta 90 \end{vmatrix} + \begin{vmatrix} \varepsilon 80 \\ \varepsilon 82 \\ \varepsilon 84 \\ \varepsilon 86 \\ \varepsilon 88 \\ \varepsilon 90 \end{vmatrix}$$

If contemporaneous correlation does not exist, single estimation produces the best, linear, unbiased estimate. However, if contemporaneous correlation does exist, SUR is a superior estimation method. A generalized least squares (GLS) procedure is used to produce a linear, unbiased estimate of β_j with minimum variance.

APPENDIX D
SPLINE REGRESSION PROCEDURE

A computer program has been created to run the spline regression for the repeat-sales technique. The data consist of last sale price (lprice), prior sale price (pprice), last sale year (lyr), prior sale year (pyr), distance from point of appreciation maximum (disthigh), and distance from appreciation minimum (distlow).

1) Create logged price ratio and (market) dummy variables:

```
gen lnpr = ln(lprice/pprice)
gen yr80 = 0
replace yr80 =  1 if lyr == 1980
replace yr80 = -1 if pyr == 1980
gen yr81 = 0
replace yr81 =  1 if lyr == 1981
replace yr81 = -1 if pyr == 1981
gen yr82 = 0
replace yr82 =  1 if lyr == 1982
replace yr82 = -1 if pyr == 1982
gen yr83 = 0
replace yr83 =  1 if lyr == 1983
replace yr83 = -1 if pyr == 1983
gen yr84 = 0
replace yr84 =  1 if lyr == 1984
replace yr84 = -1 if pyr == 1984
gen yr85 = 0
replace yr85 =  1 if lyr == 1985
replace yr85 = -1 if pyr == 1985
gen yr86 = 0
replace yr86 =  1 if lyr == 1986
replace yr86 = -1 if pyr == 1986
gen yr87 = 0
replace yr87 =  1 if lyr == 1987
replace yr87 = -1 if pyr == 1987
gen yr88 = 0
replace yr88 =  1 if lyr == 1988
replace yr88 = -1 if pyr == 1988
gen yr89 = 0
replace yr89 =  1 if lyr == 1989
replace yr89 = -1 if pyr == 1989
gen yr90 = 0
replace yr90 =  1 if lyr == 1990
replace yr90 = -1 if pyr == 1990
```

2) Run do-loop that generates submarket dummies and regression:

```
35    gen a80 = 0
36    gen a81 = 0
37    gen a82 = 0
38    gen a83 = 0
39    gen a84 = 0
40    gen a85 = 0
41    gen a86 = 0
42    gen a87 = 0
43    gen a88 = 0
44    gen a89 = 0
45    gen a90 = 0

46    gen dummy = 0
47    replace dummy = 1 if disthigh < 1.0

48    replace a80 = yr80*dummy
49    replace a81 = yr81*dummy
50    replace a82 = yr82*dummy
51    replace a83 = yr83*dummy
52    replace a84 = yr84*dummy
53    replace a85 = yr85*dummy
54    replace a86 = yr86*dummy
55    replace a87 = yr87*dummy
56    replace a88 = yr88*dummy
57    replace a89 = yr89*dummy
58    replace a90 = yr90*dummy

59    regr lnpr yr81 yr82 yr83 yr84 yr85 yr86 yr87 yr88 yr89 yr90 a81 a82
      a83 a84 a85 a86 a87 a88 a89 a90, noconstant

69    drop a80 a81 a82 a83 a84 a85 a86 a87 a88 a89 a90 dummy
```

The do-loop is run for multiple iterations. For each iteration, the increment of disthigh (in line 47) is increased by 0.1 miles. The same procedure is done for distlow. The distance at which the difference between market and submarket is greatest is determined by (the statistical significance of) individual t-statistics of the submarket dummies.

APPENDIX E
NAIVE MODEL WITH SINGLE ESTIMATION

Estimates for equation: PRICE80
Ordinary least squares (OLS) regression.

Source	SS	df	MS
Model	2.3540e+10	3	7.8468e+09
Residual	1.8243e+09	136	13414177.2
Total	2.5365e+10	139	182480121

Number of obs	=	140
F(3, 136)	=	584.96
Prob > F	=	0.0000
R-square	=	0.9281
Adj R-square	=	0.9265
Root MSE	=	3662.5

price	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
sqft	25.1585	1.546941	16.263	0.000	22.09933	28.21767
age	-220.9296	37.48812	-5.893	0.000	-295.0647	-146.7946
lot	.8618326	.1254159	6.872	0.000	.6138151	1.10985
_cons	-2528.425	2354.768	-1.074	0.285	-7185.121	2128.271

Estimates for equation: PRICE82
Ordinary least squares (OLS) regression.

Source	SS	df	MS
Model	3.6672e+10	3	1.2224e+10
Residual	3.3538e+09	136	24660576.1
Total	4.0026e+10	139	287957928

Number of obs	=	140
F(3, 136)	=	495.69
Prob > F	=	0.0000
R-square	=	0.9162
Adj R-square	=	0.9144
Root MSE	=	4965.9

price	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
sqft	32.06302	2.077704	15.432	0.000	27.95423	36.17181
age	-248.4908	47.87897	-5.190	0.000	-343.1743	-153.8072
lot	1.163501	.1635127	7.116	0.000	.8401446	1.486857
_cons	-9101.557	3129.29	-2.909	0.004	-15289.92	-2913.196

```
Estimates for equation: PRICE84
Ordinary least squares (OLS) regression.

  Source |       SS       df       MS                  Number of obs =     140
---------+------------------------------               F(  3,   136) =  605.57
   Model |  4.3738e+10     3  1.4579e+10               Prob > F      =  0.0000
Residual |  3.2742e+09   136  24075231.1               R-square      =  0.9304
---------+------------------------------               Adj R-square  =  0.9288
   Total |  4.7012e+10   139   338213946               Root MSE      =  4906.7

------------------------------------------------------------------------------
   price |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    sqft |   38.99208   2.503488     15.575   0.000      34.04128    43.94288
     age |  -285.0002    51.1085     -5.576   0.000     -386.0704     -183.93
     lot |   1.096209   .1824752      6.007   0.000      .7353528    1.457064
   _cons |  -10474.21   3371.866     -3.106   0.002     -17142.28   -3806.144
------------------------------------------------------------------------------
```

```
Estimates for equation: PRICE86
Ordinary least squares (OLS) regression.

  Source |       SS       df       MS                  Number of obs =     140
---------+------------------------------               F(  3,   136) =  851.87
   Model |  4.6953e+10     3  1.5651e+10               Prob > F      =  0.0000
Residual |  2.4987e+09   136  18372525.5               R-square      =  0.9495
---------+------------------------------               Adj R-square  =  0.9484
   Total |  4.9452e+10   139   355766366               Root MSE      =  4286.3

------------------------------------------------------------------------------
   price |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    sqft |     36.877   2.057348     17.925   0.000      32.80847    40.94553
     age |  -272.4354   43.40968     -6.276   0.000     -358.2807   -186.5901
     lot |   1.531288   .1550505      9.876   0.000      1.224666     1.83791
   _cons |  -8313.186   2833.201     -2.934   0.004     -13916.01   -2710.358
------------------------------------------------------------------------------
```

Estimates for equation: PRICE88
Ordinary least squares (OLS) regression.

Source	SS	df	MS
Model	5.3359e+10	3	1.7786e+10
Residual	4.2623e+09	136	31340242.1
Total	5.7621e+10	139	414538598

Number of obs	=	140
F(3, 136)	=	567.52
Prob > F	=	0.0000
R-square	=	0.9260
Adj R-square	=	0.9244
Root MSE	=	5598.2

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft	44.31837	2.694156	16.450	0.000	38.99051	49.64622
age	-112.3605	55.81195	-2.013	0.046	-222.732	-1.989
lot	1.612671	.200135	8.058	0.000	1.216892	2.008451
_cons	-21502.39	3639.092	-5.909	0.000	-28698.91	-14305.86

Estimates for equation: PRICE90
Ordinary least squares (OLS) regression.

Source	SS	df	MS
Model	6.1406e+10	3	2.0469e+10
Residual	4.7854e+09	136	35186917.6
Total	6.6192e+10	139	476198059

Number of obs	=	140
F(3, 136)	=	581.71
Prob > F	=	0.0000
R-square	=	0.9277
Adj R-square	=	0.9261
Root MSE	=	5931.9

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft	38.2844	2.853933	13.415	0.000	32.64058	43.92823
age	-170.962	55.8437	-3.061	0.003	-281.3963	-60.52773
lot	2.058979	.2100526	9.802	0.000	1.643588	2.474371
_cons	-14964.07	3694.981	-4.050	0.000	-22271.12	-7657.02

APPENDIX F
NAIVE MODEL WITH SIMULTANEOUS ESTIMATION

Estimates for equation: PRICE80
Generalized least squares (GLS) regression.

Observations	= 140	Weights	= ONE
Mean of LHS	= 0.3914201E+05	Std.Dev of LHS	= 0.1350852E+05
StdDev of resid	= 0.3744895E+04	Sum of squares	= 0.1907297E+10
R-squared	= 0.9225936E+00	Adjusted R-sq	= 0.920886E+00
Durbin-Watson	= 2.0689386	Autocorrelation	= -0.0344693
RHO used for GLS	= 0.0865833		

Variable	Coefficient	Std. Err	t-ratio	Prob|t|	Mean of X	Std.Dev
Constant	4487.1	2116.	2.121	0.03396		
SQFT80	21.978	1.248	17.614	0.00000	1530.5	351.72
AGE80	-282.78	34.27	-8.252	0.00000	38.414	8.9619
LOT80	878.37	106.6	8.241	0.00000	13.523	4.4084

Estimates for equation: PRICE82
Generalized least squares (GLS) regression.

Observations	= 140	Weights	= ONE
Mean of LHS	= 0.4603644E+05	Std.Dev of LHS	= 0.1696932E+05
StdDev of resid	= 0.5198864E+04	Sum of squares	= 0.3675834E+10
R-squared	= 0.9054632E+00	Adjusted R-sq	= 0.9033778E+00
Durbin-Watson	= 1.9339435	Autocorrelation	= 0.0330283
RHO used for GLS	= -0.0076121		

Variable	Coefficient	Std. Err	t-ratio	Prob|t|	Mean of X	Std.Dev.
Constant	2806.1	2593.	1.082	0.27912		
SQFT82	26.379	1.517	17.395	0.00000	1524.5	338.46
AGE82	-344.11	40.71	-8.453	0.00000	37.921	9.5096
LOT82	1191.7	125.9	9.466	0.00000	13.478	4.3766

Estimates for equation: PRICE84
Generalized least squares regression.

Observations	= 140	Weights	= ONE
Mean of LHS	= 0.5153968E+05	Std.Dev of LHS	= 0.1839059E+05
StdDev of resid	= 0.5114155E+04	Sum of squares	= 0.3557024E+10
R-squared	= 0.9221122E+00	Adjusted R-sq	= 0.9203941E+00
Durbin-Watson	= 2.0183873	Autocorrelation	= -0.0091936
RHO used for GLS	= 0.0856988		

Variable	Coefficient	Std. Err	t-ratio	Prob|t|	Mean of X	Std.Dev
Constant	1651.9	2902.	0.569	0.56921		
SQFT84	30.322	1.931	15.707	0.00000	1488.6	318.57
AGE84	-370.37	45.74	-8.098	0.00000	37.736	8.7693
LOT84	1393.5	148.3	9.394	0.00000	13.435	4.3446

Estimates for equation: PRICE86
Generalized least squares regression.

Observations	= 140	Weights	= ONE
Mean of LHS	= 0.5631911E+05	Std.Dev of LHS	= 0.1886177E+05
StdDev of resid	= 0.4418812E+04	Sum of squares	= 0.2655522E+10
R-squared	= 0.9447211E+00	Adjusted R-sq	= 0.9435017E+00
Durbin-Watson	= 1.9649866	Autocorrelation	= 0.0175067
RHO used for GLS	= 0.1426210		

Variable	Coefficient	Std. Err	t-ratio	Prob|t|	Mean of X	Std.Dev.
Constant	2072.0	2552.	0.812	0.41684		
SQFT86	30.584	1.628	18.781	0.00000	1475.5	316.48
AGE86	-351.93	40.46	-8.698	0.00000	37.379	8.9695
LOT86	1671.3	131.9	12.671	0.00000	13.324	4.1964

```
Estimates for equation: PRICE88
Generalized least squares regression.

Observations      =  140                Weights          =  ONE
Mean of LHS       =  0.6098215E+05      Std.Dev of LHS   =  0.2036022E+05
StdDev of resid   =  0.5757802E+04      Sum of squares   =  0.4508711E+10

R-squared         =  0.9194507E+00      Adjusted R-sq    =  0.9176739E+00
Durbin-Watson     =  1.9208018          Autocorrelation  =  0.0395991
RHO used for GLS  =  0.0965382

Variable  Coefficient  Std. Err  t-ratio Prob|t|   Mean of X  Std.Dev.
----------------------------------------------------------------------
Constant  -8872.4        3217.   -2.758   0.00581
SQFT88     36.333        2.058   17.657   0.00000    1473.2    303.40
AGE88     -203.52        50.50   -4.030   0.00006    37.157    9.0110
LOT88      1802.4        163.2   11.047   0.00000    13.251    4.1226
```

```
Estimates for equation: PRICE90
Generalized least squares regression.

Observations      =  140                Weights          =  ONE
Mean of LHS       =  0.6266466E+05      Std.Dev of LHS   =  0.2182196E+05
StdDev of resid   =  0.5947792E+04      Sum of squares   =  0.4811167E+10

R-squared         =  0.9251766E+00      Adjusted R-sq    =  0.9235261E+00
Durbin-Watson     =  1.9902306          Autocorrelation  =  0.0048847
RHO used for GLS     0.1471694

Variable  Coefficient  Std. Err  t-ratio Prob|t|   Mean of X  Std.Dev.
----------------------------------------------------------------------
Constant  -5780.2        3371.   -1.714   0.08645
SQFT90     31.794        2.357   13.486   0.00000    1478.5    322.25
AGE90     -219.37        52.51   -4.177   0.00003    37.400    9.4700
LOT90      2225.9        184.2   12.085   0.00000    13.317    4.3716
```

APPENDIX G
INTERACTIVE MODEL WITH SINGLE ESTIMATION

Estimates for equation: PRICE80
Ordinary least squares (OLS) regression.

Source	SS	df	MS
Model	2.3837e+11	8	2.9796e+10
Residual	1.4897e+09	132	11285378.8
Total	2.3986e+11	140	1.7133e+09

Number of obs =	140
F(8, 132) =	2640.24
Prob > F =	0.0000
R-square =	0.9938
Adj R-square =	0.9934
Root MSE =	3359.4

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft	24.07798	1.261379	19.089	0.000	21.58285	26.57311
age	-217.9145	24.07348	-9.052	0.000	-265.5342	-170.2948
lx	-546.3024	185.8211	-2.940	0.004	-913.8749	-178.7299
ly	1090.404	308.183	3.538	0.000	480.7878	1700.021
lx2	24.17981	7.52863	3.212	0.002	9.28744	39.07219
ly2	-81.21345	20.50251	-3.961	0.000	-121.7694	-40.65745
lx2y	-1.928417	.5126379	-3.762	0.000	-2.942465	-.9143682
lxy2	3.439166	.8439294	4.075	0.000	1.76979	5.108542

Estimates for equation: PRICE82
Ordinary least squares (OLS) regression.

Source	SS	df	MS
Model	3.3385e+11	8	4.1732e+10
Residual	2.8837e+09	132	21845951.9
Total	3.3674e+11	140	2.4053e+09

Number of obs =	140
F(8, 132) =	1910.26
Prob > F =	0.0000
R-square =	0.9914
Adj R-square =	0.9909
Root MSE =	4674.0

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft	27.69811	1.693468	16.356	0.000	24.34826	31.04795
age	-333.625	33.76213	-9.882	0.000	-400.4098	-266.8401
lx	-966.0692	256.572	-3.765	0.000	-1473.594	-458.5444
ly	1829.804	430.3064	4.252	0.000	978.615	2680.992
lx2	41.93702	10.38598	4.038	0.000	21.39252	62.48151
ly2	-127.9116	28.90292	-4.426	0.000	-185.0844	-70.73879
lx2y	-3.175686	.7137249	-4.449	0.000	-4.587505	-1.763868
lxy2	5.386727	1.18835	4.533	0.000	3.036052	7.737401

Estimates for equation: PRICE84
Ordinary least squares (OLS) regression.

Source	SS	df	MS
Model	4.1599e+11	8	5.1998e+10
Residual	2.9125e+09	132	22064153.1
Total	4.1890e+11	140	2.9921e+09

Number of obs	= 140
F(8, 132)	= 2356.69
Prob > F	= 0.0000
R-square	= 0.9930
Adj R-square	= 0.9926
Root MSE	= 4697.2

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft	31.18561	2.05336	15.188	0.000	27.12386	35.24736
age	-385.1117	35.8829	-10.732	0.000	-456.0916	-314.1318
lx	-956.0176	266.306	-3.590	0.000	-1482.797	-429.238
ly	1795.145	449.081	3.997	0.000	906.8185	2683.472
lx2	42.46633	10.82194	3.924	0.000	21.05945	63.8732
ly2	-122.9602	29.99816	-4.099	0.000	-182.2996	-63.62091
lx2y	-3.195662	.7454879	-4.287	0.000	-4.670311	-1.721014
lxy2	5.253959	1.235059	4.254	0.000	2.810891	7.697027

Estimates for equation: PRICE86
Ordinary least squares (OLS) regression.

Source	SS	df	MS
Model	4.9133e+11	8	6.1416e+10
Residual	2.1841e+09	132	16546212.7
Total	4.9351e+11	140	3.5251e+09

Number of obs	= 140
F(8, 132)	= 3711.77
Prob > F	= 0.0000
R-square	= 0.9956
Adj R-square	= 0.9953
Root MSE	= 4067.7

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft	32.00891	1.666486	19.207	0.000	28.71244	35.30538
age	-391.3471	32.52526	-12.032	0.000	-455.6853	-327.009
lx	-729.6795	236.5278	-3.085	0.002	-1197.555	-261.8042
ly	1420.61	399.9107	3.552	0.000	629.5471	2211.673
lx2	34.00906	9.629595	3.532	0.000	14.96077	53.05735
ly2	-92.92281	27.09196	-3.430	0.000	-146.5134	-39.33224
lx2y	-2.575662	.6685466	-3.853	0.000	-3.898113	-1.253211
lxy2	4.00718	1.11872	3.582	0.000	1.794241	6.22012

Estimates for equation: PRICE88
Ordinary least squares (OLS) regression.

Source	SS	df	MS
Model	5.7388e+11	8	7.1735e+10
Residual	4.3748e+09	132	33142101.0
Total	5.7826e+11	140	4.1304e+09

Number of obs	= 140
F(8, 132)	= 2164.47
Prob > F	= 0.0000
R-square	= 0.9924
Adj R-square	= 0.9920
Root MSE	= 5756.9

price	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
sqft	33.38288	2.376855	14.045	0.000	28.68123	38.08454
age	-396.0799	45.427	-8.719	0.000	-485.939	-306.2208
lx	-874.6428	332.7011	-2.629	0.010	-1532.758	-216.5272
ly	1757.464	559.2833	3.142	0.002	651.1462	2863.782
lx2	39.36362	13.506	2.915	0.004	12.64741	66.07983
ly2	-114.4702	37.48219	-3.054	0.003	-188.6137	-40.32675
lx2y	-3.000065	.9294883	-3.228	0.002	-4.838684	-1.161445
lxy2	4.771301	1.544148	3.090	0.002	1.716824	7.825777

Estimates for equation: PRICE90
Ordinary least squares (OLS) regression.

Source	SS	df	MS
Model	6.1193e+11	8	7.6491e+10
Residual	4.0203e+09	132	30456710.7
Total	6.1595e+11	140	4.3997e+09

Number of obs	= 140
F(8, 132)	= 2511.48
Prob > F	= 0.0000
R-square	= 0.9935
Adj R-square	= 0.9931
Root MSE	= 5518.8

price	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
sqft	31.2539	2.216352	14.102	0.000	26.86973	35.63806
age	-393.9619	43.2524	-9.108	0.000	-479.5194	-308.4043
lx	-776.0801	312.7203	-2.482	0.014	-1394.672	-157.4884
ly	1611.346	525.0574	3.069	0.003	572.7305	2649.961
lx2	37.37298	12.69001	2.945	0.004	12.27088	62.47507
ly2	-104.9904	35.40538	-2.965	0.004	-175.0257	-34.95503
lx2y	-2.911431	.8767774	-3.321	0.001	-4.645784	-1.177079
lxy2	4.473703	1.461916	3.060	0.003	1.581888	7.365517

APPENDIX H
INTERACTIVE MODEL WITH SIMULTANEOUS ESTIMATION

Estimates for equation: PRICE80
Generalized least squares regression.

Observations	= 140	Weights	= ONE
Mean of LHS	= 0.3914201E+05	Std.Dev of LHS	= 0.1350852E+05
StdDev of resid	= 0.3345748E+04	Sum of squares	= 0.1477612E+10
R-squared	= 0.9382148E+00	Adjusted R-sq	= 0.9349384E+00
Durbin-Watson	= 2.0306799	Autocorrelation	= -0.0153399
RHO used for GLS	= 0.1056754		

Variable	Coefficient	Std. Err	t-ratio	Prob|t|	Mean of X	Std.Dev.
SQFT80	23.574	1.053	22.397	0.00000	1530.5	351.72
AGE80	-186.30	21.13	-8.815	0.00000	38.414	8.9619
LX80	-168.63	145.0	-1.163	0.24490	328.19	132.63
LY80	478.11	238.7	2.003	0.04516	189.93	76.046
LX280	9.0441	5.876	1.539	0.12380	8170.6	4124.8
LY280	-43.551	15.92	-2.735	0.00623	2781.1	1555.0
LX2Y80	-0.93949	0.4013	-2.341	0.01923	117570	71045.
LXY280	1.9130	0.6615	2.892	0.00383	69171.	45347.

Estimates for equation: PRICE82
Generalized least squares regression.

Observations	= 140	Weights	= ONE
Mean of LHS	= 0.4603644E+05	Std.Dev of LHS	= 0.1696932E+05
StdDev of resid	= 0.4629486E+04	Sum of squares	= 0.2829043E+10
R-squared	= 0.9250365E+00	Adjusted R-sq	= 0.9210612E+00
Durbin-Watson	= 2.1125196	Autocorrelation	= -0.0562598
RHO used for GLS	= 0.0739243		

Variable	Coefficient	Std. Err	t-ratio	Prob|t|	Mean of X	Std.Dev.
SQFT82	26.482	1.235	21.447	0.00000	1524.5	338.46
AGE82	-266.91	27.43	-9.730	0.00000	37.921	9.5096
LX82	-569.65	191.0	-2.983	0.00285	327.15	131.87
LY82	1206.1	318.8	3.783	0.00016	188.99	74.227
LX282	26.249	7.747	3.388	0.00070	8144.5	4111.8
LY282	-92.232	21.64	-4.262	0.00002	2762.7	1515.3
LX2Y82	-2.1938	0.5368	-4.087	0.00004	117030	70404.
LXY282	3.9552	0.8989	4.400	0.00001	68742.	44580.

Estimates for equation: PRICE84
Generalized least squares regression.

Observations = 140		Weights = ONE	
Mean of LHS = 0.5153968E+05		Std.Dev of LHS = 0.1839059E+05	
StdDev of resid = 0.4693314E+04		Sum of squares = 0.2907590E+10	
R-squared = 0.9344034E+00		Adjusted R-sq = 0.9309248E+00	
Durbin-Watson = 2.0218867		Autocorrelation = -0.0109433	
RHO used for GLS = 0.0982745			

Variable	Coefficient	Std. Err	t-ratio	Prob|t|	Mean of X	Std.Dev.
SQFT84	30.024	1.579	19.013	0.00000	1488.6	318.57
AGE84	-317.08	30.02	-10.56	0.00000	37.736	8.7693
LX84	-499.74	199.8	-2.501	0.01240	326.29	130.91
LY84	1056.9	333.5	3.169	0.00153	187.99	71.158
LX284	24.951	8.186	3.048	0.00230	8128.3	4085.7
LY284	-80.874	22.46	-3.600	0.00032	2741.0	1445.6
LX2Y84	-2.1065	0.5685	-3.705	0.00021	116490	68663.
LXY284	3.6259	0.9403	3.856	0.00012	68221	42850.

Estimates for equation: PRICE86
Generalized least squares regression.

Observations = 140		Weights = ONE	
Mean of LHS = 0.5631911E+05		Std.Dev of LHS = 0.1886177E+05	
StdDev of resid = 0.3966479E+04		Sum of squares = 0.2076750E+10	
R-squared = 0.9554591E+00		Adjusted R-squared= 0.9530971E+00	
Durbin-Watson = 1.9967455		Autocorrelation = 0.0016273	
RHO used for GLS = 0.2111593			

Variable	Coefficient	Std. Err	t-ratio	Prob|t|	Mean of X	Std.Dev.
SQFT86	31.377	1.329	23.614	0.00000	1475.5	316.48
AGE86	-335.39	27.15	-12.35	0.00000	37.379	8.9695
LX86	-547.31	193.2	-2.833	0.00462	323.85	128.75
LY86	1128.5	319.5	3.532	0.00041	186.83	71.638
LX286	27.429	7.884	3.479	0.00050	8070.9	4039.9
LY286	-79.501	21.42	-3.711	0.00021	2731.3	1474.2
LX2Y86	-2.2152	0.5443	-4.070	0.00005	115900	68721.
LXY286	3.5491	0.8979	3.953	0.00008	68014	43428.

Estimates for equation: PRICE88
Generalized least squares regression.

Observations	= 140	Weights	= ONE
Mean of LHS	= 0.6098215E+05	Std.Dev of LHS	= 0.2036022E+05
StdDev of resid	= 0.5684023E+04	Sum of squares	= 0.4264671E+10
R-squared	= 0.9215018E+00	Adjusted R-squared	= 0.9173390E+00
Durbin-Watson	= 2.0394957	Autocorrelation	= -0.0197478
RHO used for GLS	= 0.1516012		

Variable	Coefficient	Std. Err	t-ratio	Prob|t|	Mean of X	Std.Dev.
SQFT88	32.480	1.782	18.231	0.00000	1473.2	303.40
AGE88	-325.14	37.46	-8.679	0.00000	37.157	9.0110
LX88	-398.33	262.9	-1.515	0.12973	321.59	126.16
LY88	986.22	434.7	2.269	0.02327	186.06	71.668
LX288	20.907	10.67	1.960	0.05003	8005.2	3970.0
LY288	-70.243	29.02	-2.420	0.01550	2725.1	1486.7
LX2Y88	-1.8442	0.7305	-2.525	0.01158	115160	68486.
LXY288	3.0442	1.207	2.522	0.01167	67775.	43643.

Estimates for equation: PRICE90
Generalized least squares regression.

Observations	= 140	Weights	= ONE
Mean of LHS	= 0.6266466E+05	Std.Dev of LHS	= 0.2182196E+05
StdDev of resid	= 0.5453471E+04	Sum of squares	= 0.3925725E+10
R-squared	= 0.9370970E+00	Adjusted R-squared	= 0.9337612E+00
Durbin-Watson	= 1.9810445	Autocorrelation	= 0.0094777
RHO used for GLS	= 0.1111796		

Variable	Coefficient	Std. Err	t-ratio	Prob|t|	Mean of X	Std.Dev.
SQFT90	30.931	1.806	17.123	0.00000	1478.5	322.25
AGE90	-317.11	37.09	-8.549	0.00000	37.400	9.4700
LX90	-406.13	261.2	-1.555	0.11992	323.25	130.00
LY90	1031.1	435.9	2.365	0.01801	186.70	73.094
LX290	23.088	10.59	2.179	0.02930	8046.9	4027.8
LY290	-75.235	29.41	-2.558	0.01053	2729.2	1485.7
LX2Y90	-2.0700	0.7332	-2.823	0.00475	115590	68688.
LXY290	3.3354	1.222	2.729	0.00635	67894.	43554.

APPENDIX I
APPRECIATION MODELS

Equation 1: Square Footage

Estimates for equation: APPR
Ordinary least squares regression.

Source	SS	df	MS
Model	.000796239	1	.000796239
Residual	.004858753	138	.000035208
Total	.005654992	139	.000040683

Number of obs	=	140
F(1, 138)	=	22.62
Prob > F	=	0.0000
R-square	=	0.1408
Adj R-square	=	0.1346
Root MSE	=	.00593

appr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft	-.0000078	.0000016	-4.756	0.000	-.000011	-.0000045
_cons	.0646419	.0024974	25.883	0.000	.0597037	.0695801

Equation 2: Age

Estimates for equation: APPR
Ordinary least squares regression.

Source	SS	df	MS
Model	.000625507	1	.000625507
Residual	.005029485	138	.000036446
Total	.005654992	139	.000040683

Number of obs	=	140
F(1, 138)	=	17.16
Prob > F	=	0.0000
R-square	=	0.1106
Adj R-square	=	0.1042
Root MSE	=	.00604

appr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0002407	.0000581	4.143	0.000	.0001258	.0003556
_cons	.0439209	.0022518	19.505	0.000	.0394683	.0483734

Equation 3: Lot Size

Estimates for equation: APPR
Ordinary least squares regression.

Source	SS	df	MS
Model	.001077214	1	.001077214
Residual	.004577778	138	.000033172
Total	.005654992	139	.000040683

Number of obs	= 140
F(1, 138)	= 32.47
Prob > F	= 0.0000
R-square	= 0.1905
Adj R-square	= 0.1846
Root MSE	= .00576

appr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lot	-.0006627	.0001163	-5.699	0.000	-.0008927	-.0004328
_cons	.0618792	.0016312	37.934	0.000	.0586538	.0651046

Equation 4: Square Footage, Age, and Lot Size

Estimates for equation: APPR
Ordinary least squares regression.

Source	SS	df	MS
Model	.001281086	3	.000427029
Residual	.004373906	136	.000032161
Total	.005654992	139	.000040683

Number of obs	= 140
F(3, 136)	= 13.28
Prob > F	= 0.0000
R-square	= 0.2265
Adj R-square	= 0.2095
Root MSE	= .00567

appr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft	.0000005	.0000030	0.170	0.865	-.0000055	.0000065
age	.0001476	.0000586	2.517	0.013	.0000317	.0002635
lot	-.0005848	.0002236	-2.616	0.010	-.0010270	-.0001427
_cons	.0544913	.0039129	13.926	0.000	.0467534	.0622292

Equation 5: Price

Estimates for equation: APPR
Ordinary least squares regression.

Source	SS	df	MS
Model	.000974373	1	.000974373
Residual	.004680619	138	.000033918
Total	.005654992	139	.000040683

Number of obs =	140
F(1, 138) =	28.73
Prob > F =	0.0000
R-square =	0.1723
Adj R-square =	0.1663
Root MSE =	.00582

appr	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
actp80	-.0000002	.00000004	-5.360	0.000	-.0000002	-.00000012
_cons	.0606788	.00151360	40.089	0.000	.0576860	.06367170

Equation 6: Polynomial Expansion of (X,Y) Coordinates

Estimates for equation: APPR
Ordinary least squares regression.

Source	SS	df	MS
Model	.005633974	5	.001126795
Residual	.000021019	134	1.5686e-07
Total	.005654992	139	.000040683

Number of obs =	140
F(5, 134) =	7183.55
Prob > F =	0.0000
R-square =	0.9963
Adj R-square =	0.9961
Root MSE =	.0004

appr	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
x	-.0029955	.0001611	-18.598	0.000	-.0033141	-.0026770
y	.0044855	.0000733	61.225	0.000	.0043406	.0046304
x2	.0001794	.0000044	40.572	0.000	.0001707	.0001881
x2y	-.0000105	.0000001	-70.692	0.000	-.0000108	-.0000102
xy2	.0000014	.0000001	7.677	0.000	.0000010	.0000017
_cons	.0372827	.0018221	20.461	0.000	.0336789	.0408865

APPENDIX J
REPEAT-SALES MODELS

Spline regression for: Northeast
Radial Area: 2.5 miles
(predicted negative abnormal appreciation)

Estimates for repeat-sales equation: lnpr
Ordinary least squares regression.

Source	SS	df	MS
Model	194.557496	20	9.72787478
Residual	98.8168293	3978	.024840832
Total	293.374325	3998	.073380271

Number of obs	=	3998
F(20, 3978)	=	391.61
Prob > F	=	0.0000
R-square	=	0.6632
Adj R-square	=	0.6615
Root MSE	=	.15761

lnpr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
yr81	.0873677	.0095803	9.119	0.000	.0685848	.1061505
yr82	.1442588	.0092155	15.654	0.000	.1261912	.1623264
yr83	.2040763	.0085370	23.905	0.000	.1873391	.2208136
yr84	.2707677	.0087044	31.107	0.000	.2537023	.2878332
yr85	.3242259	.0084684	38.287	0.000	.3076231	.3408286
yr86	.3857562	.0082434	46.796	0.000	.3695946	.4019179
yr87	.4236192	.0085542	49.522	0.000	.4068482	.4403902
yr88	.4516835	.0087432	51.661	0.000	.4345420	.4688250
yr89	.4711132	.0091291	51.606	0.000	.4532151	.4890113
yr90	.4860392	.0088383	54.993	0.000	.4687112	.5033672
a81	-.0239935	.0326771	-0.734	0.463	-.0880589	.0400719
a82	-.0329716	.0298553	-1.104	0.269	-.0915047	.0255615
a83	-.0041118	.0288702	-0.142	0.887	-.0607136	.0524900
a84	-.0320856	.0293899	-1.092	0.275	-.0897063	.0255351
a85	-.0623629	.0296535	-2.103	0.036	-.1205004	-.0042254
a86	-.0661533	.0277776	-2.382	0.017	-.1206130	-.0116935
a87	-.0907208	.0292220	-3.105	0.002	-.1480122	-.0334293
a88	-.1005559	.0305903	-3.287	0.001	-.1605300	-.0405818
a89	-.1074609	.0297293	-3.615	0.000	-.1657470	-.0491749
a90	-.095776	.0315633	-3.034	0.002	-.1576578	-.0338941

Spline regression for: Northwest
Radial Area: 3.5 miles
(predicted positive abnormal appreciation)

Estimates for repeat-sales equation: lnpr
Ordinary least squares regression.

Source	SS	df	MS
Model	194.14049	20	9.7070245
Residual	99.2338349	3978	.02494566
Total	293.374325	3998	.073380271

Number of obs	= 3998
F(20, 3978)	= 389.13
Prob > F	= 0.0000
R-square	= 0.6618
Adj R-square	= 0.6600
Root MSE	= .15794

lnpr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
yr81	.0850951	.0095056	8.952	0.000	.0664588	.1037314
yr82	.1408195	.0090679	15.530	0.000	.1230414	.1585976
yr83	.1996303	.0084210	23.706	0.000	.1831205	.2161402
yr84	.266552	.0086028	30.984	0.000	.2496857	.2834183
yr85	.3156139	.0083710	37.703	0.000	.2992021	.3320257
yr86	.3747102	.0081496	45.979	0.000	.3587325	.3906880
yr87	.4133248	.0084459	48.938	0.000	.3967660	.4298836
yr88	.4369203	.0086881	50.289	0.000	.4198866	.4539539
yr89	.4582962	.0090260	50.775	0.000	.4406002	.4759923
yr90	.4769421	.0087934	54.239	0.000	.4597021	.4941821
a81	.0278783	.0367651	0.758	0.448	-.0442018	.0999584
a82	.0228317	.0368016	0.620	0.535	-.0493201	.0949834
a83	.0646941	.0348268	1.858	0.063	-.0035860	.1329741
a84	.0472541	.0349200	1.353	0.176	-.0212087	.1157169
a85	.070106	.0351375	1.995	0.046	.0012169	.1389952
a86	.0968098	.0329693	2.936	0.003	.0321716	.1614481
a87	.0658732	.0351438	1.874	0.061	-.0030284	.1347748
a88	.1042177	.0345235	3.019	0.003	.0365323	.1719032
a89	.0716731	.0351065	2.042	0.041	.0028446	.1405016
a90	.0498854	.0346776	1.439	0.150	-.0181022	.117873

REFERENCE LIST

Alonso, W., *Location and Land Use*, Harvard University Press: Cambridge, MA, (1964)

Anderson, R.J., and T.D. Croker, "Air Pollution and Residential Property Values," *Urban Studies*, Volume 8, pp 171-180, (1971)

Archer W.R., D.H. Gatzlaff, and D.C. Ling, "Measuring the Importance of Location on House Price Appreciation," forthcoming in *Journal of Urban Economics*, (1995)

Asabere, P.K. and B. Harvey, "Factors Influencing the Value of Urban Land: Evidence from Halifax-Dartmouth, Canada," *AREUEA Journal*, Volume 13, pp 361-377, (1985)

Bailey, M.J., R.F. Muth, and H.O. Norse, "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association*, Volume 58, pp 933-942, (1963)

Balkin, S. and J.F. McDonald, "The Market for Streetcrime: An Economic Analysis of Victims-Offender Interaction," *Journal of Urban Economics*, Volume 10, pp 390-405, (1981)

Ball, M.J., "Recent Empirical Work on the Determinants of Relative House Prices," *Urban Studies*, Volume 10, pp 213-231, (1973)

Berry, B.J.L., "Ghetto Expansion and Single-Family Housing Prices: Chicago, 1968-1972," *Journal of Urban Economics*, Volume 3, pp 397-423, (1976)

Berry, B.J.L., and R. Bednarz, "The Disbenefits of Neighborhood and Environment to Urban Property," *Geographic Humanism, Analysis and Social Action*, Ann Arbor, MI, Geographical Publication #17, pp 111-148, (1977)

Brigham, E.F., "The Determinants of Residential Land Values," *Land Economics*, Volume 41, pp 325-334, (1965)

Case, K.E., "The Market for Single-Family Homes in the Boston Area," *New England Economic Review*, (May/June), pp 38-48, (1986)

Case, K.E., and R.J. Shiller, "Prices of Single-Family Homes since 1970: New Indexes for Four Cities," *New England Economic Review*, (Sept/Oct), pp 45-56, (1987)

Case, K.E., and R.J. Shiller, "The Efficiency of the Market for Single-Family Homes," *The American Economic Review*, Volume 79, pp 125-137, (1989)

Case, K.E., and R.J. Shiller, "A Decade of Boom and Bust in the Prices of Single-Family Homes; Boston and Los Angeles, 1983 to 1993," *New England Economic Review*, March/April, pp 40-51, (1994)

Chicoine, D.L., "Farmland Values at the Urban Fringe: An Analysis of Sale Prices," *Land Economics*, Volume 57, pp 353-362, (1981)

Chorley, R.J., and P. Haggett, "Trend Surface Mapping in Geographical Research," *Transactions and Papers of the Institute of British Geographers*, Publication 37, pp 47-67, (1965)

Clapp, J.M., and C. Giaccotto, "Estimating Price Indices for Residential Properties: A Comparison of Repeat Sales and Assessed Value Methods," *Journal of the American Statistical Association*, Volume 87, pp 300-306, (1991)

Cliff, A. and J. Ord, *Spatial Autocorrelation*, Pion: London, UK, (1973)

deLeeuw, F. and R. J. Struyk, *The Web of Urban Housing*, The Urban Institute: Washington, D. C., (1975)

Dubin, Robin A., "Spatial Autocorrelation and Neighborhood Quality," *Regional Science and Urban Economics*, Volume 22, pp 433-452, (1992)

Fik, T. J., E. J. Malecki, and R. G. Amey, "Trouble in Paradise? Employment Trends and Forecasts for a Service-Oriented Economy," *Economic Development Quarterly*, Volume 7, Number 4, pp 358-372, (1993)

Gatzlaff, D.H., and D.C. Ling, "Measuring Changes in Local House Prices: An Empirical Investigation of Alternative Methodologies," *Journal of Urban Economics*, Volume 35, pp 221-244, (1994)

Getis, A., "Second-Order Analysis of Point Patterns: The Case of Chicago as a Multicenter Urban Region," *The Professional Geographer*, Volume 35, pp 73-80, (1983)

Gillard, Q., "The Effect of Environmental Amenities on House Values: The Example of a View Lot," *Professional Geographer*, Volume 33, pp 216-220, (1981)

Goodman, A.C., "Hedonic Prices, Price Indices, and Housing Markets," *Journal of Urban Economics*, Volume 5, pp 471-484, (1978)

Green, D.L., "Urban Subcenters: Recent Trends in Urban Spatial Structure," *Growth and Change*, Volume 11, pp 29-40, (1980)

Hall, R.E., "The Measurement of Quality Change From Vintage Price Data," in Griliches, Z. (ed.) *Price Index and Quality Change*, Cambridge: Cambridge, MA, (1971)

Halvorsen, R. and H.O. Pollakowski "Choice of Functional Form for Hedonic Price Equations," *Journal of Urban Economics*, Volume 10, pp 37-49, (1981)

Haurin, D.R., and P.H. Hendershott, "House Price Indices: Issues and Results," *AREUEA Journal*, Volume 19, pp 259-269, (1991)

Harrison, D., and D. Rubinfeld, "Hedonic Housing Prices and the Demand for Clean Air," *Journal of Environmental Economics and Management*, Volume 5, pp 81-102, (1978)

Heikkila, E., P. Gordon, J.I. Kim, R.B. Peiser, and H.W. Richardson, "What Happened to the CBD-Distance Gradient?: Land Values in a Polycentric City", *Environment and Planning A*, Vol. 21, pp 221-232, (1989)

Hembd, J., and C.L. Infanger, "Application of Trend Surface Analysis to a Rural Housing Market," *Land Economics*, Volume 57, pp 303-322, (1981)

Hendershott, P.H., and T. Thibodeau, "The Relationship Between Median and Constant Quality House Prices: Implications for Setting FHA Loan Limits," *AREUEA Journal*, Volume 18, pp 323-334, (1990)

Henderson, J.V., *Economic Theory and the Cities*, Academic Press: New York, NY, (1977)

Jackson, J. R. "Intraurban Variation in the Price of Housing," *Journal of Urban Economics*, Volume 6, pp 464-479, (1979)

Johnson, M. S. and W. R. Ragas "CBD Land Values and Multiple Externalities," *Land Economics*, Volume 63, pp 337-347, (1987)

Judge, G. G., *The Theory and Practice of Econometrics*, Wiley: New York, NY, (1985)

Kain, J.F. and J.M. Quigley, "Measuring the Value of Housing Quality," *Journal of the American Statistical Association*, Volume 65, pp 532-548, (1970)

Keil, K.A. and R.T. Carson, "An Examination of Systematic Differences in the Appreciation of Individual Housing Units," *Journal of Real Estate Research*, Volume 5, pp 301-318, (1990)

Knos, D. S., "The Distribution of Land Values in Topeka, Kansas," in *Spatial Analysis*, edited by B. J. L. Berry and D. F. Marble, Prentice Hall: Englewood Cliffs, NJ, (1968)

Krumbein, W.C. and F.A. Graybill, *An Introduction to Statistical Modeling in Geology*, McGraw-Hill: New York, NY, (1965)

Lancaster, K., "A New Approach to Consumer Theory", *Journal of Political Economy*, Volume 74, pp 132-157, (1966)

Li, M.M., and Brown, H.J., "Micro-Neighborhood Externalities and Hedonic Housing Prices," *Land Economics*, Volume 56, pp 125-141, (1980)

Linneman, A., "Some Empirical Results on the Nature of the Hedonic Price Function for the Urban Housing Market," *Journal of Urban Economics*, Volume 8, pp 47-68, (1980)

Linneman, A., "The Demand for Residence Site Characteristics," *Journal of Urban Economics*, Volume 9, pp 129-148, (1981)

Little, J.T., "Residential Preferences, Neighborhood Filtering, and Neighborhood Change," *Journal of Urban Economics*, Volume 3, pp 68-81, (1976)

Marshall, A., *Principles of Economics*, 8th ed., Macmillan and Co.: London, UK, (1920)

Mills, E.S., *Studies in the Structure of the Urban Economy*, John Hopkins Press: Baltimore, MD, (1972)

Muth, R.F., *Cities and Housing*, Chicago University Press: Chicago, IL, (1969)

Muth, R. F., "Numerical Solution of Urban Residential Land Use Models," *Journal of Urban Economics*, Volume 2, pp 307-332, (1975)

Palmquist, R.B., "Alternative Techniques for Developing Real Estate Price Indices," *Review of Economics and Statistics*, Volume 62, pp 442-480, (1980)

Parker, C., "Trend Surface and the Spatio-Temporal Analysis of Residential Land Use Intensity and Household Housing Expenditure," *Land Economics*, Volume 57, pp 323-337, (1981)

Rachlis, M.B. and A.M.J. Yezer, "Urban Location and Housing Price Appreciation," *Papers of the Regional Science Association*, Volume 57, pp 155-164, (1985)

Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy, Volume 82, pp 34-55, (1974)

Smith, A., *An Inquiry into the Nature and Causes of the Wealth of Nations*, Clarendon Press: Oxford, UK, (1776)

Thayer, M., H. Albers, and M. Rahmatian, " The Benefits of Reducing Exposure to Waste Disposal Sites: A Hedonic Housing Value Approach," *Journal of Real Estate Research*, Volume 7, pp 265-282, (1992)

von Bohm-Bawerk, E., *The Positive Theory of Capital*, (1888), translated by Smart, W., Macmillan and Co.: London, UK, (1891)

von Thünen, J.H., *Den Isolierte Staadt*, (1826), translated: The Isolated State by Wartenberg, C.M., Permagon Press: New York, NY, (1966)

Waddell, Paul, Brian J.L. Berry and I. Hoch, "Housing Price Gradients: The Intersection of Space and Built Form", *Geographical Analysis*, Vol. 25, pp 5-19, (1993a)

Waddell, P., B.J.L. Berry, and I. Hoch, "Residential Property Values in a Multinodal Urban Area: New Evidence on the Implicit Price of Location," *Journal of Real Estate Finance and Economics*, Volume 7, pp 117-141, (1993b)

Weicher, J., and R. Zerbst, "The Externalities of Neighborhood Parks: An Empirical Investigation," *Land Economics*, Volume 49, pp 99-105, (1973)

Zellner, A., *Readings in Economic Statistics and Econometrics*, Little, Brown: Boston, (1968)

BIOGRAPHICAL SKETCH

Greg T. Smersh is the son of Dr. Jerome F. Smersh and Martha E. Smersh and was raised in Seattle, Washington. He graduated from the University of Washington in 1981 with a Bachelor of Science in Building Construction and from Western Washington University in 1984 with a masters in business administration. From 1978 to 1983, Greg worked as a general contractor for Smersh Brothers Construction and from 1984 to 1990 as a securities analyst for the firms of Shearson Lehman Brothers and Cable, Howse, and Ragen. He is currently completing his Ph.D. in geography at the University of Florida and conducting research that focuses on the spatial aspects of both residential and commercial real estate.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Timothy J. Fik, Chairman
Associate Professor of Geography

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Edward J. Malecki
Professor of Geography

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Peter R. Waylen
Associate Professor of Geography

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John R. Dunkle
John R. Dunkle
Professor of Geography

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

David C. Ling
David C. Ling
Associate Professor of Finance, Insurance, and Real Estate

This dissertation was submitted to the Graduate Faculty of the Department of Geography in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1995

Dean, Graduate School